The Astrophysical Journal, 890:122 (24pp), 2020 February 20 https://doi.org/10.3847/1538-4357/ab67bc

Absolute Calibration Strategies for the Hydrogen Epoch of Reionization Array and Their Impact on the 21 cm Power Spectrum

Nicholas S. Kern[1], Joshua S. Dillon[1,21], Aaron R. Parsons[1], Christopher L. Carilli[2], Gianni Bernardi[3,4,5], Zara Abdurashidova[1], James E. Aguirre[6], Paul Alexander[7], Zaki S. Ali[1], Yanga Balfour[5], Adam P. Beardsley[8], Tashalee S. Billings[6], Judd D. Bowman[8], Richard F. Bradley[9], Philip Bull[10], Jacob Burba[11], Steven Carey[7], Carina Cheng[1], David R. DeBoer[1], Matt Dexter[1], Eloy de Lera Acedo[7], John Ely[7], Aaron Ewall-Wice[12], Nicolas Fagnoni[7], Randall Fritz[5], Steve R. Furlanetto[13], Kingsley Gale-Sides[7], Brian Glendenning[2], Deepthi Gorthi[1], Bradley Greig[14], Jasper Grobbelaar[5], Ziyaad Halday[5], Bryna J. Hazelton[15,16], Jacqueline N. Hewitt[12], Jack Hickish[1], Daniel C. Jacobs[8], Austin Julius[5], Joshua Kerrigan[11], Piyanat Kittiwisit[8], Saul A. Kohn[6], Matthew Kolopanis[8], Adam Lanman[11], Paul La Plante[6], Telalo Lekalake[5], Adrian Liu[1,17,22], David MacMahon[1], Lourence Malan[5], Cresshim Malgas[5], Matthys Maree[5], Zachary E. Martinot[6], Eunice Matsetela[5], Andrei Mesinger[18], Mathakane Molewa[5], Miguel F. Morales[15], Tshegofalang Mosiane[5], Steven G. Murray[8], Abraham R. Neben[12], Bojan Nikolic[7], Chuneeta D. Nunhokee[1], Nipanjana Patra[1], Samantha Pieterse[5], Jonathan C. Pober[11], Nima Razavi-Ghods[7], Jon Ringuette[15], James Robnett[2], Kathryn Rosie[5], Peter Sims[11], Craig Smith[5], Angelo Syce[5], Nithyanandan Thyagarajan[2,8], Peter K. G. Williams[19,20], and Haoxuan Zheng[12]

[1] Department of Astronomy, University of California, Berkeley, CA, USA; nkern@berkeley.edu
[2] National Radio Astronomy Observatory, Socorro, NM, USA
[3] Department of Physics and Electronics, Rhodes University, P.O. Box 94, Grahamstown, 6140, South Africa
[4] INAF–Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy
[5] SKA-SA, Cape Town, South Africa
[6] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA, USA
[7] Cavendish Astrophysics, University of Cambridge, Cambridge, UK
[8] School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA
[9] National Radio Astronomy Observatory, Charlottesville, VA, USA
[10] School of Physics & Astronomy, Queen Mary University of London, London, UK
[11] Department of Physics, Brown University, Providence, RI, USA
[12] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA
[13] Department of Physics and Astronomy, University of California, Los Angeles, CA, USA
[14] School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
[15] Department of Physics, University of Washington, Seattle, WA, USA
[16] eScience Institute, University of Washington, Seattle, WA, USA
[17] Department of Physics and McGill Space Institute, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada
[18] Scuola Normale Superiore, I-56126 Pisa, PI, Italy
[19] Center for Astrophysics, Harvard & Smithsonian, 60 Garden Street, Cambridge, MA, USA
[20] American Astronomical Society, 1667 K Street NW, Suite 800, Washington, DC 20006, USA
Received 2019 October 28; revised 2019 December 27; accepted 2020 January 2; published 2020 February 19

Abstract

We discuss absolute calibration strategies for Phase I of the Hydrogen Epoch of Reionization Array (HERA), which aims to measure the cosmological 21 cm signal from the Epoch of Reionization. HERA is a drift-scan array with a 10° wide field of view, meaning bright, well-characterized point-source transits are scarce. This, combined with HERA's redundant sampling of the *uv* plane and the modest angular resolution of the Phase I instrument, make traditional sky-based and self-calibration techniques difficult to implement with high dynamic range. Nonetheless, in this work, we demonstrate calibration for HERA using point-source catalogs and electromagnetic simulations of its primary beam. We show that unmodeled diffuse flux and instrumental contaminants can corrupt the gain solutions and present a gain-smoothing approach for mitigating their impact on the 21 cm power spectrum. We also demonstrate a hybrid sky and redundant calibration scheme and compare it to pure sky-based calibration, showing only a marginal improvement to the gain solutions at intermediate delay scales. Our work suggests that the HERA Phase I system can be well calibrated for a foreground avoidance power spectrum estimator by applying direction-independent gains with a small set of degrees of freedom across the frequency and time axes.

Unified Astronomy Thesaurus concepts: Cosmology (343); Reionization (1383); Astronomy data analysis (1858)

1. Introduction

The Hydrogen Epoch of Reionization Array[23] (HERA; DeBoer et al. 2017) is a targeted, radio interferometric experiment that aims to measure the cosmological 21 cm spin-flip emission from primordial hydrogen in the intergalactic medium (IGM) at Cosmic Dawn . One of the last frontiers of cosmology and high-redshift astrophysics, the Cosmic Dawn marks the era when the first stars, black holes, and galaxies formed and interacted with the surrounding IGM. Eventually, these sources heated and reionized the majority of the neutral hydrogen in the IGM in an event known as the Epoch of Reionization (EoR). A number of questions remain about when and how the Cosmic Dawn and EoR occurred that are crucial to our broader understanding of galaxy and large-scale structure

[21] NSF Astronomy and Astrophysics Postdoctoral Fellow.
[22] Hubble Fellow.
[23] http://reionization.org/

formation. For reviews, see Morales & Wyithe (2010), Mesinger (2016), and Liu & Shaw (2019).

One of the only direct probes of the IGM throughout the entirety of Cosmic Dawn is neutral hydrogen's 21 cm transition, which at redshifts of $z \sim 10$ appears in the low-frequency radio band around 150 MHz. Over the past decade, first-generation 21 cm EoR experiments like the Donald C. Backer Precision Array for Probing the EoR (PAPER; Parsons et al. 2014; Jacobs et al. 2015; Cheng et al. 2018; Kolopanis et al. 2019), Murchison Widefield Array (MWA; Tingay et al. 2013; Dillon et al. 2014; Beardsley et al. 2016; Ewall-Wice et al. 2016a; Barry et al. 2019b; Li et al. 2019), Low Frequency Array (LOFAR; van Haarlem et al. 2013; Patil et al. 2017; Gehlot et al. 2019), Giant Metre Wave Radio Telescope (GMRT; Paciga et al. 2013), and Long Wavelength Array (LWA; Eastwood et al. 2019) have set increasingly stringent limits on the Cosmic Dawn 21 cm power spectrum. Meanwhile, global signal experiments have placed constraints on the 21 cm monopole (Bernardi et al. 2016; Singh et al. 2017), with a reported first detection of the signal at Cosmic Dawn from the Experiment to Detect the Global EoR Signature (EDGES; Bowman et al. 2018). The 21 cm experiments face the challenge of separating out the weak cosmological signal from galactic and extragalactic foreground emission that is many orders of magnitude brighter. However, the 21 cm signal is expected to be highly spectrally variant due to inhomogeneities in the density, ionization state, and temperature of the IGM along the line of sight, while nonthermal foreground emission is expected to be spectrally smooth. This provides a means for separating foreground emission from the desired cosmological signal. However, even small instrumental effects can distort these foregrounds and contaminate the region in Fourier space nominally occupied only by the EoR signal and thermal noise, known as the EoR window (Morales et al. 2012). High dynamic range instrumental gain calibration is therefore critical to 21 cm science.

Per-antenna gain calibration is the task of solving for a single complex number per antenna and feed polarization (as a function of both time and frequency) that best satisfies the antenna-based calibration equation for a visibility V_{ij} defined between antennas i and j,

$$V_{ij}^{\rm measured}(\nu, t) = V_{ij}^{\rm true}(\nu, t) g_i(\nu, t) g_j^*(\nu, t), \quad (1)$$

where $V_{ij}^{\rm measured}$ is the raw data, $V_{ij}^{\rm true}$ is the true visibility that would be measured by an uncorrupted instrument, and g_i and g_j are the instrumental gains for antennas i and j, respectively (Hamaker et al. 1996). Recent work has shown how incomplete models in sky-based calibration (Barry et al. 2016; Ewall-Wice et al. 2017; Byrne et al. 2019) and nonredundancies in redundant calibration (Joseph et al. 2018; Orosz et al. 2019) can lead to gain calibration errors that contaminate the EoR window. Foreground and instrument simulations for HERA indicate that the fiducial EoR signal at $k \sim 0.2\,h\,{\rm Mpc}^{-1}$ is expected to be roughly 10^5 times weaker than the peak foreground amplitude at $k \sim 0\,h\,{\rm Mpc}^{-1}$ in the visibility (Thyagarajan et al. 2016). Because gain calibration is multiplicative in frequency, it can equivalently be thought of as a convolution in delay space, the Fourier dual of frequency. This means that each antenna's gain kernel, or the gain's footprint in delay space, must be nominally suppressed by at least 5 orders of magnitude at delay scales of $\tau \gtrsim 400$ ns (400 ns equals $k_\parallel = 0.2\ h\ {\rm Mpc}^{-1}$ at $z = 10$ or $\nu \sim 130$ MHz for the 21 cm line). In this case, we have chosen to represent the gains as direction-independent, which is the component of gains we are concerned with in this work, although much work has been devoted to direction-dependent gain calibration (e.g., Bhatnagar et al. 2008; Intema 2014).

HERA was deployed in two stages, Phase I and Phase II. Phase I observed from 2017 to 2018 while only a section of the array was built and used front-end signal chains from the PAPER experiment. Phase II is currently under construction toward a build-out of 350 antennas and will be equipped with completely new front-end hardware. The work in this paper uses only Phase I observations (Section 2). HERA is a drift-scan array, meaning it is built into the ground and cannot physically point its antennas on the sky. With its 10° degree field of view (FoV), the number of bright and well-characterized point sources that transit on any given night is limited. Furthermore, the highly redundant *uv* sampling and relatively short baselines of the HERA Phase I configuration make implementing self-calibration to high dynamic ranges difficult. Nonetheless, we outline a strategy for sky-based calibration of HERA Phase I using point sources from the MWA's GLEAM catalog (Hurley-Walker et al. 2017) and electromagnetic simulations of HERA's primary beam (Fagnoni et al. 2019). We show that this does a fairly good job of bringing the data in line with the adopted model and use it to characterize the frequency and time stability of the gains. Importantly, we also show that performing antenna-based calibration in the presence of non-antenna-based systematics can contaminate systematic-free visibilities. We discuss the impact this has on the data and the 21 cm power spectrum and demonstrate gain-smoothing procedures to mitigate this and other gain errors introduced in the process of calibration.

Redundant calibration has been hailed as a powerful alternative calibration strategy for 21 cm experiments that sidesteps some of the requirements of sky-based calibration (Liu et al. 2010; Zheng et al. 2014). However, redundant calibration still needs a sky model to pin down certain degenerate parameters it cannot solve for (Dillon et al. 2018; Joseph et al. 2018; Li et al. 2018; Byrne et al. 2019). In this work, we explore hybrid-redundant absolute calibration strategies using the `hera_cal` package.[24] Applying them to HERA Phase I, we show that redundant calibration seems to mitigate some errors associated with sky-based calibration; however, it also has its own set of uncertainties due to inherent nonredundancies that need to be mitigated. For low delay modes in the gains, we find that redundant and sky calibration yield very similar results.

In this work, we use the term *absolute calibration* to refer to the components of the full antenna-based gains that are constant across the array (note that these are still frequency-dependent). One example of this is the average antenna gain amplitude, which sets the overall flux scale of the data. Indeed, these are exactly the terms that are degenerate in redundant calibration. In sky-based calibration, these terms are automatically solved for, which can therefore be thought of as a form of absolute calibration.

The structure of this paper is as follows. In Section 2 we detail the observations used in this analysis. In Section 3 we describe our methodology for sky-based calibration of HERA.

[24] https://github.com/HERA-Team/hera_cal

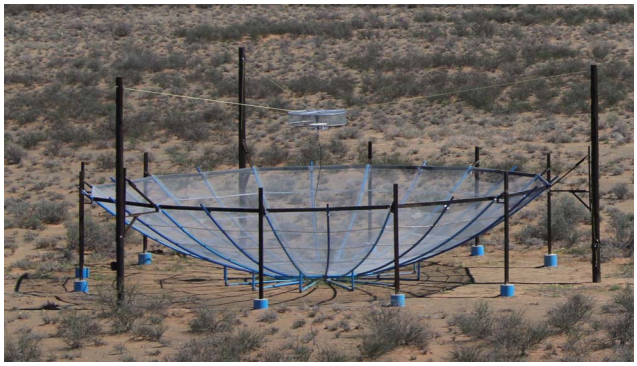

Figure 1. Single HERA antenna in the field with a cross-dipole feed surrounded by a cage hoisted to the antenna's focal point. Image courtesy of Kathryn Rosie.

In Section 4 we characterize the time and frequency stability of the gain solutions. In Section 5 we synthesize redundant and absolute calibration and compare them to traditional sky-based calibration. In Section 6 we calibrate the data and investigate foreground contamination in the power spectrum, and in Section 7 we summarize our results.

2. Observations

The data used in this work were taken with the HERA Phase I instrument (DeBoer et al. 2017) in a 56-element configuration on 2017 December 10 (Julian Date 2,458,098). HERA is located in the Karoo Desert, South Africa, at the South African Karoo Radio Astronomy Reserve. Data were taken in drift-scan mode for roughly 12 hr night^{-1} starting at 5 pm South African Standard Time, of which roughly 9 hr are deemed good-quality data when the Sun is below the horizon.

The Phase I instrument repurposed many of the older PAPER experiment components, including its signal chains, correlator, feeds, and front-end modules (FEMs), and attached them to newly designed HERA antennas. The HERA antenna (Figure 1) is a 14 m dish with an optimized version of the dual linear polarization PAPER feed and FEM hoisted 4.9 m to its focal height. The optimized feed and dish were designed to minimize reflections within the antenna and thus limit excess chromaticity induced by the signal chain (Ewall-Wice et al. 2016b; Neben et al. 2016; Thyagarajan et al. 2016; Patra et al. 2018). From the FEM, which houses an initial stage of amplification, the analog chain consists of a 150 m coaxial cable connected to a node unit in the field where the signals are fed through a postamplification stage (PAM) and a filtering stage. From there, the signals travel through another 20 m coaxial cable to a container where they are digitized, Fourier-transformed, and cross-multiplied with all other antenna and linear polarization streams. Additional observational parameters are detailed in Table 1.

Not all of the PAPER signal chains could be salvaged for the HERA Phase I instrument. As a temporary stopgap, additional FEMs, cables, and PAMs were manufactured for Phase I data collection. We refer to the new set of signal chains as "Type 1" and the old set of signal chains as "Type 2," which are colored blue and red in Figure 2, respectively. The transmission properties of the signal chains are studied in more detail in Kern et al. (2019a). For more details on the HERA Phase I signal chain and electronics, we refer the reader to Parsons et al. (2010), DeBoer et al. (2017), and Fagnoni et al. (2019).

Table 1
HERA Observation Parameters

Parameter	Value
Array configuration	Phase I
Number of antennas	56
Array coordinates	−30.°7 S, 21.°4 E
Observing mode	Drift-scan
Correlator integration	10.7 s
Frequency range	100–200 MHz
Channel width	97.65 kHz
Dish diameter	14 m
Feed type	Dual-polarization *X* and *Y* dipoles
Visibility polarizations	*XX*, *XY*, *YX*, *YY*
Shortest/longest baseline	14.6/139.3 m
Observation dates	2017 December 10

Note. For the 2017–2018 observation, the HERA correlator used the convention that the *X* dipole points east–west, while the *Y* dipole points north–south. This is not the standard Hamaker & Bregman (1996) definition, which assumes the opposite.

Before calibration, the data are preprocessed with part of the HERA analysis pipeline. Specifically, faulty antennas are identified and flagged at a quality metrics stage (crosses in Figure 2), and radio frequency interference (RFI) is excised from the data using median filtering and a watershed algorithm (Kerrigan et al. 2019). The data are written to disk in the Miriad file format postcorrelation and then converted to `UVFITS` using the `pyuvdata` software (Hazelton et al. 2017) and imported to Common Astronomy Software Applications (CASA) measurement sets via CASA's `importuvfits` task.

3. Sky-based Calibration

Standard sky-based calibration is typically done by choosing a bright, well-characterized point source for the model visibilities. This is made difficult for HERA because it is a drift-scan array, meaning it cannot be pointed to an arbitrary location on the sky. Furthermore, the larger collecting area provided by a dish, as opposed to a lone dipole, means HERA's primary beam response is more compact on the sky compared to other experiments like PAPER or the MWA: at 150 MHz, HERA's primary beam FWHM is roughly 10°, compared to roughly 45° for the PAPER experiment. This means that the number of bright, well-characterized radio sources that transit our FoV is low. In fact, not a single point source within 5° of HERA's decl. exceeds 20 Jy in flux density in the cold part of the radio sky (far from the galactic plane). Implementing self-calibration to a high dynamic range is also difficult for HERA, given its highly redundant sampling of the *uv* plane, making HERA's narrowband grating lobes very severe. This is compounded by the poor angular resolution of the Phase I instrument, making it quickly confusion noise-limited (Figure 2). Redundant calibration somewhat skirts the problem of an inadequate sky model, and indeed, exploiting the power of redundant calibration was a motivating factor behind HERA's redundant design (Dillon & Parsons 2016). However, redundant calibration operates only within a specific subspace of the full antenna-based calibration equations, meaning a model of the sky is still fundamentally needed to fill in the few remaining degenerate modes (Liu et al. 2010; Zheng et al. 2014; Dillon et al. 2018; Li et al. 2018; Byrne et al. 2019; J.

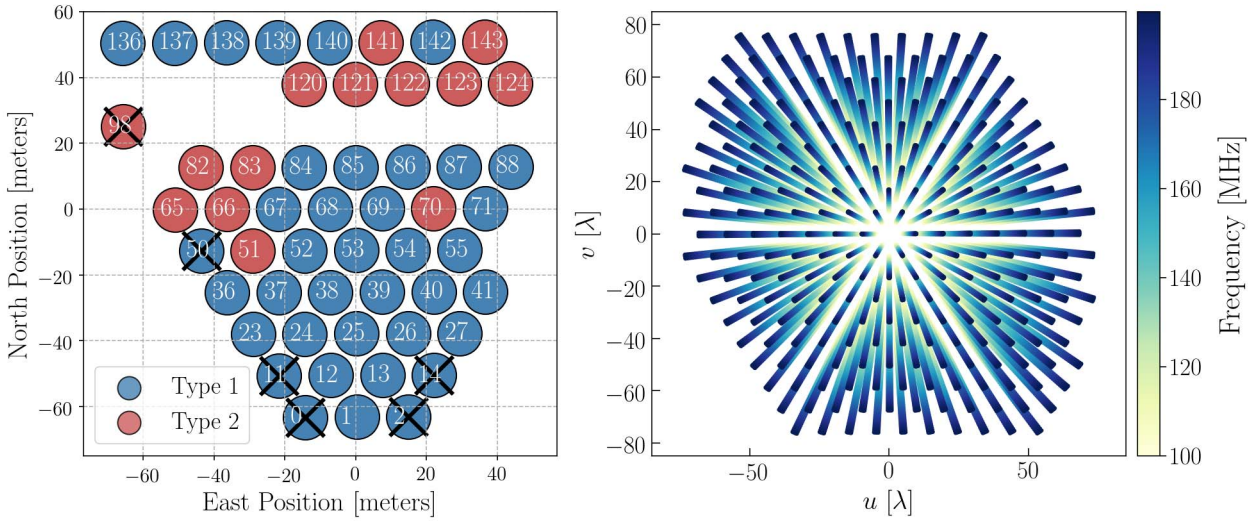

Figure 2. Left: HERA Phase I array layout with 56 connected antennas and 50 operational antennas. Antennas determined to be problematic are marked with crosses. Right: corresponding *uv* sampling of the array over a 10 minute time window and a frequency range of 100–200 MHz, highlighting HERA's highly redundant *uv* sampling. The color gradient represents independent *uv* samples throughout the total bandwidth.

Dillon et al. 2020, in preparation). We discuss this in more detail for HERA in Section 5.

For power spectrum estimators that do not attempt to subtract the dominant foreground emission in the data (at the expense of losing low k modes), the stringent requirement of high dynamic range source modeling is relaxed because we are not interested in recovering modes inherently occupied by foreground emission. Hybrid techniques also exist, which try to reap the benefits of both foreground removal and avoidance (Kerrigan et al. 2018). For foreground avoidance estimators, a path toward achieving deep, noise-limited power spectrum integrations at intermediate spatial modes of $k \gtrsim 0.2\,h\,\mathrm{Mpc}^{-1}$ with a calibration derived from the sky may be possible even with the challenges faced by the HERA Phase I instrument. In this section, we describe a sky-based calibration strategy for HERA using custom pipelines for calibration and imaging[25] built around the CASA (McMullin et al. 2007) package. We start by discussing the construction of our flux density model and then describe our calibration methodology and its validation via imaging and source extraction.

3.1. Building a Sky Model

Our ideal model for sky-based calibration would involve a single bright point source located at the pointing center of the FoV. Because HERA is a drift-scan array, this means our ideal calibrator would be located at $\delta \sim -30.^{\circ}7$ and would transit zenith at some point in the night. Ideally, this calibrator would be so bright that other off-axis point sources or diffuse emission would contribute a vastly subdominant component of the measured visibilities. Unfortunately, this is not the case for HERA, so we are forced to make compromises. Figure 3 is a map of radio foregrounds at 150 MHz from the Global Sky Model (de Oliveira-Costa et al. 2008) and shows the HERA stripe (white dashed lines), which denotes the track of the FWHM of HERA's primary beam (10° at 150 MHz). We see that the HERA stripe covers a fairly small part of the sky, demonstrating how limited we are in the amount of sky available for identifying bright calibrators.

[25] https://github.com/HERA-Team/casa_imaging

To select the best calibration field given our limitations, we can identify some key criteria that a good field should satisfy. The first criterion is that the field should have most of its radio emission contained in the main lobe of the primary beam. Off-axis sources located in the far side lobes of the primary beam are troublesome because primary beam side lobes are hard to model accurately. One work-around is to peel these sources from the visibilities before calibration (e.g., Hurley-Walker et al. 2017; Eastwood et al. 2019), but that requires one to image them at a fine frequency resolution to capture primary beam chromaticity and with high dynamic range, which, as stated, is challenging for HERA Phase I. Additionally, we want our direction-independent calibration to be representative of the instrument response at zenith, because that is where most of the measured EoR signal comes from. Said another way, we do not want our direction-independent calibration to soak up structure from direction-dependent effects introduced by off-axis sources. One example of this is diffuse emission coming from the plane of the galaxy, which extends across the entire FoV when it transits.

The second criterion for a good calibration field is that it should have sources that are well characterized at the observing frequencies. Furthermore, it should have a relatively bright source very close to the FoV pointing center so that we can confirm via imaging that our calibration at zenith yields a good match to the input model. Such a source can also be useful for empirically characterizing the primary beam response with drift-scan source tracks (Pober et al. 2012; Eastwood et al. 2018).

Recently, the MWA constructed the GLEAM point-source catalog (Hurley-Walker et al. 2017) from a deep, low-frequency survey spanning the southern hemisphere, overlapping with the HERA stripe. We searched the GLEAM catalog for all point sources within $2.^{\circ}5$ of $\delta = -30.^{\circ}7$ with a flux density above 15 Jy at 150 MHz located in the cold part of the radio sky (local sidereal time (LST) < 6 hr). We find three such sources in the GLEAM catalog: J0024–2929 at 0 hr LST, J0200–3053 at 2 hr LST, and J0455–3006 at 5 hr LST. Their positions, flux densities, and spectral indices are reported in Table 2. Jacobs (2016) performed a similar exercise with the

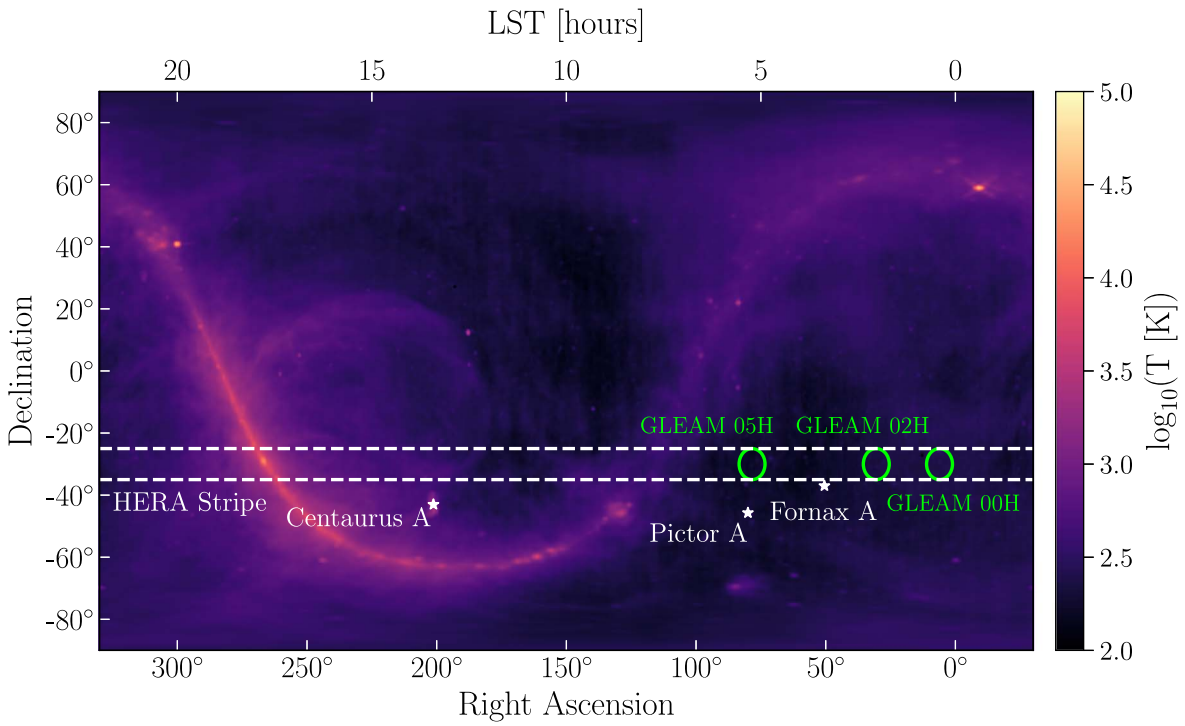

Figure 3. Radio sky at 150 MHz from the GSM (de Oliveira-Costa et al. 2008), showing the bright galactic and extragalactic foregrounds that stand in the way of cosmological 21 cm experiments. The HERA stripe is shown by white dashed lines centered at HERA's decl. of $-30.^{\circ}7$ with a width of 10°, which is the FWHM of the primary beam at 150 MHz. The three fields identified as ideal calibration fields are shown by green circles, and some bright extended sources in the vicinity are marked by stars.

Table 2
HERA Calibrator Candidates from GLEAM

Name	R.A. (J2000)	Decl. (J2000)	$S_{\rm peak}$	$S_{\rm int}$	α
J0024–2929	6.126	−29.48	16.45	16.10	−0.867
J0200–3053	30.05	−30.89	19.50	17.95	−0.863
J0455–3006	73.81	−30.11	16.34	17.11	−0.781

Note. All GLEAM (Hurley-Walker et al. 2017) sources above 15 Jy, with LST < 6 hr and $-33.2 < \delta < -28.2$. Equatorial coordinates are in degrees, flux densities are in Jy at 151 MHz, and α is the spectral index anchored at 151 MHz.

TGSS ADR catalog (Intema et al. 2017). They also found J0200–3053 as a possible calibrator but did not identify the other two sources we quote from the GLEAM catalog. For the shared source, the quoted values between the GLEAM and TGSS ADR catalogs agree to within 15%, which is roughly in line with the overall accuracy of the survey flux scales. The green circles in Figure 3 are centered on each of these three calibration fields and have diameters equal to the 10° FWHM of the HERA primary beam at 150 MHz. Stars mark the location of nearby bright extended sources like Pictor A and Fornax A.

Even though an ∼20 Jy primary calibrator source exists at the pointing center of each field, they themselves make up only a fraction of the total flux density measured by the instrument at those LSTs. For short baselines, the dominant sky component is diffuse galactic emission, while longer baselines are dominated by point sources spread across the FoV. Although models of the diffuse galactic emission exist (de Oliveira-Costa et al. 2008; Zheng et al. 2017), they are only accurate at ∼15% and furthermore extend across the entire FoV, filling the hard-to-model side lobes. At the moment, we only use point sources in our flux density model and cut short baselines (<40 m) that have significant amounts of diffuse foreground emission. Our starting model for each field is made up of all GLEAM point sources down to 0.1 Jy in flux density extending 20° in radius from the pointing center, which typically results in ∼10,000 sources in the flux density model. We take the GLEAM-reported flux density of each point source at 151 MHz and their spectral indexes and insert them into a CASA component list. All sources are assumed to be unpolarized, and their fluxes are inserted purely as Stokes *I*. For GLEAM sources without a spectral index, we take the reported flux density of the source at 122, 130, 143, 151, 158, 166, and 174 MHz and fit our own spectral index. After constructing a component list with all of the relevant GLEAM sources, we make a 1024-channel spectral cube image of the component list with the CASA `Image.modify` task, matching the channelization of HERA data, and export it to `FITS` format. The image has a pixel resolution of 300″, which is six times smaller than the synthesized beam FWHM of ∼0.5 °.

Note that the GLEAM catalog does not include bright extended sources like Fornax A and Pictor A. As shown in Figure 3, the calibration fields are chosen such that these sources are heavily attenuated by the primary beam, but even so, these sources can be seen at the level of a few Jy for the 2 and 5 hr fields, for example. Fornax A and Pictor A can be included in the component list model for the GLEAM-02H and GLEAM-05H fields, respectively, by adopting point-source models with spectral indices informed by recent low-frequency studies (Jacobs et al. 2013; McKinley et al. 2015). Although these sources have a non-zero angular extent to them, for

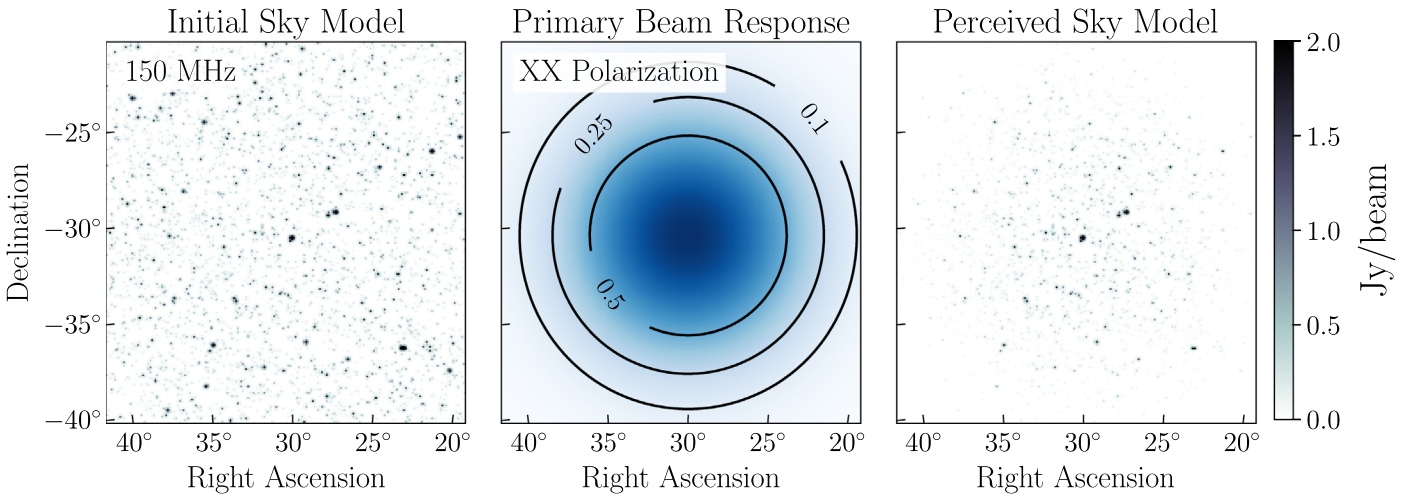

Figure 4. Construction of the GLEAM-02H field sky model for calibration at 150 MHz. Each frequency channel in the model is constructed independently in the same manner. Left: all GLEAM point sources in Stokes I polarization above 0.1 Jy within 20° of the pointing center. In this figure, the point sources have been convolved with a narrow 2D Gaussian merely for visual clarity. Middle: peak-normalized primary beam response for the *XX* instrumental linear polarization at 150 MHz (Fagnoni et al. 2019). Right: Stokes *I* model multiplied by the *XX* primary beam response yielding a perceived flux density model that is then converted into visibilities for calibration.

HERA Phase I angular resolutions, a point-source model is a fair approximation.

Next, we incorporate the effects of the direction and frequency-dependent antenna primary beam response to create a perceived flux density model. We use an electromagnetic simulation of the HERA primary beam from Fagnoni et al. (2019), which includes effects from the dish and feed. That work also explored the effects of mutual coupling on the primary beam response given an element embedded in the array, finding second-order effects on the beam response near the horizon at a level of 10^{-2} in power. Empirical studies by Kern et al. (2019a) find similar levels of mutual coupling in the data and present postcalibration methods for mitigating their effects. In this work, we only use the Fagnoni et al. (2019) beam model of the antenna and feed, and we defer using the embedded element pattern in calibration for future work. Each linear dipole in the feed, *X* and *Y*, is assigned its own beam model, where one is simply a 90° rotation of the other. The beams are peak-normalized at boresight independently at each frequency, and we then multiply the beam response at each pixel on the sky separately for the *X* and *Y* dipoles. This results in two spectral cubes, one each for the *XX* and *YY* instrumental visibility polarizations, which constitute our perceived model. In this work, we do not construct models for the cross-polarized *XY* and *YX* visibilities, as we will not perform polarization calibration, although this can be done with polarized beam models (Martinot et al. 2018). Figure 4 demonstrates this for the GLEAM-02H field in *XX* instrumental polarization, showing the initial sources (left), the *XX* primary beam response (or the squared *X*-dipole response) at 150 MHz (middle), and the product of the two (right). Lastly, the model cubes are transformed from the image to the *uv* domain via CASA's `ft` task and inserted into the model column of the measurement sets for calibration.

3.2. Calibration

Next, we will describe our approach for deriving complex, direction-independent antenna gains with CASA. For simplicity, we will focus our discussion specifically on the GLEAM-02H field, but calibration on any other field would follow the same procedures outlined below. As noted, the data are first processed for faulty antennas and RFI flagging by the HERA analysis pipeline. We then take 5 minutes of drift-scan data centered at the transit of the primary calibrator, apply a fringe-stop phasing to the transit LST, and time-average the data. Averaging 5 minutes of data allows us to increase the signal-to-noise ratio (S/N) of the derived gains and is still a fairly short time interval compared to the FWHM primary beam crossing time (at 150 MHz) of ∼46 minutes, at which point sky source decorrelation will begin to be a problem. Due to the inherent stability of the drift-scan observing mode, we do not expect the gains to vary substantially over such short timescales (although see Section 4.2 for higher-order effects).

Before proceeding with calibration, we enact a minimum baseline cut such that all baselines shorter than 40 m (∼20λ) are excluded, leaving 65% of the visibilities for calibration. HERA's shortest baselines are most sensitive to the diffuse galactic emission that is not included in our point-source model. After experimenting with various baseline cuts, we find a 40 m cut to be a good compromise between keeping as many data as possible for maximal gain S/N and eliminating diffuse foreground flux in the data that is not included in our model.

Our process for deriving antenna gains uses a series of standard routines in CASA. Before each calibration step, we apply all previous calibration steps to the data on the fly. The final calibration is then simply the product of all steps in our calibration chain. We start by performing delay calibration using the `gaincal` task, which is done to calibrate out the cable delay of each antenna. Next, we perform mean-phase and mean-amplitude calibration (which consists of two numbers for each antenna polarization across the entire bandwidth), also using the `gaincal` task. This removes any residual phase offset after delay calibration and sets the overall flux scale of the data. Up to now, all calibration steps are smooth across frequency and therefore do not contain significant spectral structure. Finally, we derive complex antenna bandpasses using the `bandpass` task, which solves each frequency channel independently from all others. This last step has the possibility of introducing an arbitrary amount of spectral structure into the

gain solutions and therefore deserves closer attention, which we revisit in Section 4.

In this work, we do not make any attempt to correct for effects due to the ionosphere. This is less of a concern given the higher frequency range of 100–200 MHz, as well as the limited angular resolution of the array and the fact that observations are only taken at night when the Sun is below the horizon, leading to calmer ionospheric conditions. We also do not attempt to calibrate the relative phase between dipole polarizations in this work, which is difficult due to the dearth of bright polarized sky sources (Lenc et al. 2017; Moore et al. 2017), although this can still be partially constrained if we assert that the Stokes V visibilities are consistent with thermal noise (Kohn et al. 2016). This is less of a concern because in this work, we are mostly interested in the parallel-hand (i.e., XX and YY) dipole and Stokes I data products, which are not as sensitive to this term as the Stokes U and V data products. While previous work has shown that ionospheric leakage of point-source foregrounds can, in principle, be significant (Nunhokee et al. 2017), ionospheric-induced leakage terms have also been shown to average down night to night (Martinot et al. 2018). As we will show in Section 3.3, the amount of intrinsic polarization leakage observed in the data is quite small, even without performing any kind of polarization calibration. Future HERA observations that (i) extend below 100 MHz or (ii) are interested in polarized data products will need to revisit these topics. For an investigation into direction-dependent effects and polarization leakage from the HERA-19 commissioning array, see Kohn et al. (2018).

3.3. Imaging

To test the fidelity of the calibration, we make multi-frequency synthesis (MFS) images of the calibrated data, the calibration model, and their residual visibility as a visual assessment of their agreement. The MFS images use 5 minutes of data and a 60 MHz bandwidth spanning 120–180 MHz. All images are made from only the baselines involved in the calibration ($|b| > 40$ m), employ robust weighting with `robust` $= -1$, and adopt the Högbom deconvolution algorithm (Högbom 1974) using the `tclean` task. All images are CLEANed independently down to a threshold of 0.5 Jy in the polarization they are imaged in. CLEAN masks are initially used around the brightest sources, and then the CLEAN mask is opened up to the entire field. We produce images in instrumental XX and YY polarization and also pseudo-Stokes I, Q, U, and V polarization.

The HERA array is not perfectly coplanar, which will introduce artifacts into wide-field images made with CASA. This can be mitigated with W-projection (Cornwell et al. 2008); however, given the FoV and modest angular resolution of the Phase I array, we do not expect non-coplanar effects to generate an appreciable amount of error. Therefore, we do not perform W-projection in the process of imaging, which also reduces its overall computational cost.

Figure 5 shows the GLEAM-02H field in XX polarization and images of its calibrated data (left), model (middle), and their residual visibility (right). The size of the synthesized beam is shown in the lower left corner. We see good agreement between the data and model down to a few percent. The residual image appears noise-like in the main lobe, but further away from the pointing center, we can begin to correlate point sources in the data with point sources in the residual. This is a result of an improper perceived flux density model (either with the inherent source fluxes or, more likely, the adopted primary beam response). This will introduce spectrally dependent errors into the gain solutions at some level (Barry et al. 2016; Ewall-Wice et al. 2017), which we explore in the following section. This can be partially mitigated by self-calibration or redundant calibration, although redundant calibration still suffers from this effect to some degree (Byrne et al. 2019).

We also make images of the pseudo-Stokes visibilities as a diagnostic tool. The pseudo-Stokes visibilities (Hamaker et al. 1996) are a linear sum of the linear polarization visibilities, defined as

$$\begin{pmatrix} V_I \\ V_Q \\ V_U \\ V_V \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & i & 0 \end{pmatrix}\begin{pmatrix} V_{XX} \\ V_{XY} \\ V_{YX} \\ V_{YY} \end{pmatrix}. \quad (2)$$

Note that these are not true Stokes parameters, which are only properly defined in the image plane, but can be thought of as approximations to the true Stokes visibility one would form by Fourier transforming the true Stokes parameter from the image plane to the uv plane. In the limit that the instrumental (direction-dependent) Mueller matrix is the identity matrix, then the pseudo-Stokes visibility defined in Equation (2) is identical to the true Stokes visibility. In practice, we do not expect this to be the case, except for possibly near the pointing center in the image, where, after having performed direction-independent calibration, we hope direction-dependent terms are minimal.

We do not expect appreciable levels of polarized sources in the GLEAM-02H field. For a recent study by the MWA, see Lenc et al. (2017). Given an ideal telescope with no instrumental leakage, we would therefore expect the pseudo-Q, U, and V visibilities to look noise-like. However, we know that the primary beam response at a given point on the sky for the X and Y dipoles is not the same at low zenith angles, which will by itself cause polarization leakage of observed off-axis sources into Stokes Q (Moore et al. 2017). Furthermore, we have not attempted to calibrate feed D-terms (Hamaker et al. 1996) or the unconstrained relative X–Y phase parameter leftover after Stokes I calibration (Sault et al. 1996). We also know from previous studies that mutual coupling exists at a nonnegligible level (Fagnoni et al. 2019; Kern et al. 2019a), which is, in principle, a direction-dependent term in the Mueller matrix. This means that we (i) wholly expect that images formed from pseudo-Stokes visibilities will not necessarily be representative of the true Stokes parameters in the image plane, except for maybe near the pointing center, and (ii) should observe nonnegligible amounts of polarization leakage from Stokes $I \rightarrow$ Stokes Q, U, and V. To properly make true Stokes parameters one would image each of the linear dipole visibilities and perform direction-dependent corrections in the image plane before adding them in a similar manner as Equation (2). At the moment, we defer this to future works that more carefully consider polarization calibration.

Figure 6 shows MFS images of the GLEAM-02H field in pseudo-Stokes I, Q, U, and V (left to right). The first thing to note is that the observed leakage of Stokes I to Q, U, and V is on the order of a few percent, which is quite low given that we did not apply a polarization or direction-dependent calibration. Looking at the pseudo-Stokes Q image, we can see the effects

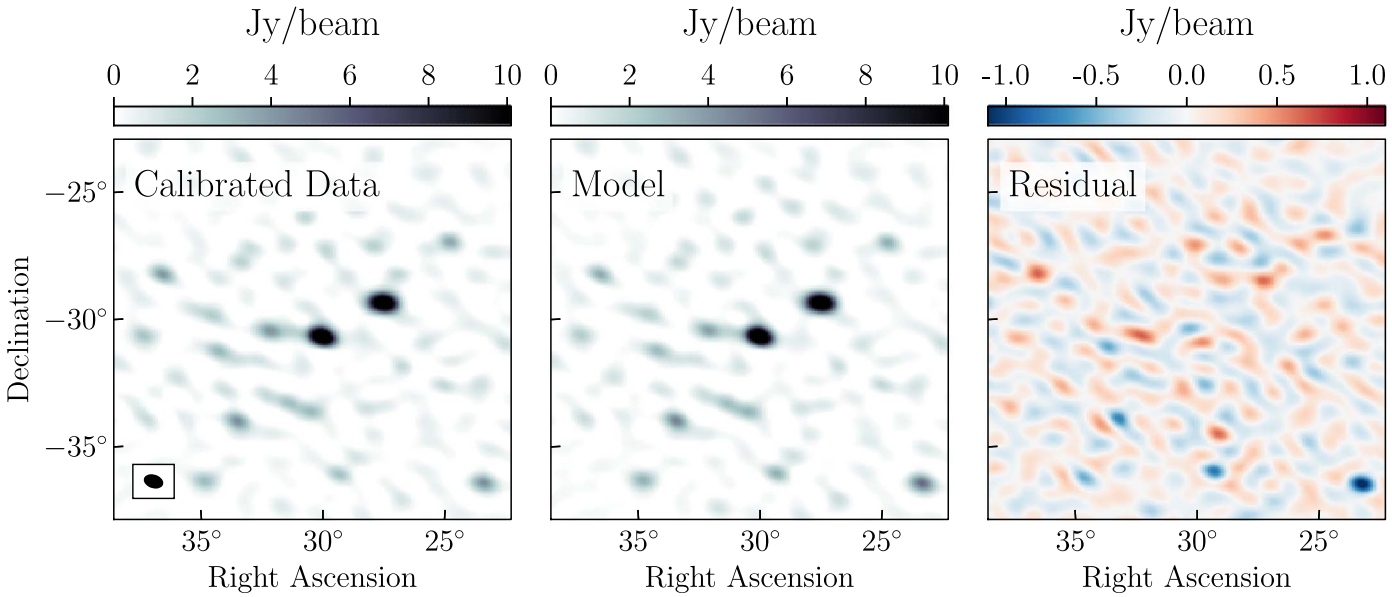

Figure 5. The MFS image of the GLEAM-02H field in *XX* polarization spanning 120–180 MHz of the calibrated (left), model (center), and residual (right) visibilities. Each image is CLEANed with the same parameters down to 0.5 Jy, with the restoring beam shown in the lower left corner. The model and calibrated data show good agreement in the main lobe of the primary beam. At larger zenith angles, the residual image shows evidence for miscalibration, likely due to primary beam errors.

of primary beam asymmetry between *X* and *Y* dipoles: without a primary beam correction (which is not applied here), the asymmetry will cause leakage of $I \rightarrow Q$ (Moore et al. 2017), which is exacerbated the more discrepant the primary beam responses are at a given point on the sky. Although nearly azimuthally symmetric, the *X*-dipole beam is elongated along the north–south direction, while the *Y* dipole is elongated along the east–west direction (Martinot et al. 2018; Fagnoni et al. 2019). This means we might expect the relative amplitude of the *X* and *Y* beams to attain a better match in the corner of our images and would therefore expect to see more $I \rightarrow Q$ leakage in a quadrupolar pattern on the sky. Indeed, this is observed in the pseudo-Stokes *Q* image to some degree (Figure 6).

The pseudo-Stokes *U* and *V* images also exhibit interesting behaviors, in particular the sources in the pseudo-Stokes *U* image that are clearly correlated with true Stokes *I* sources, as well as the rumble in the pseudo-Stokes *V* image that seems to be concentrated near the main lobe. This could be due to polarization leakage stemming from the uncalibrated *X*–*Y* phase term; however, further work is needed to identify its exact cause.

Having shown that our calibration does a fairly good job bringing our data in line with our model (Figure 5) and that, even without polarization calibration, polarization leakage is observed at a few percent (Figure 6), we should also show that our derived bandpass is an accurate solution as a function of frequency. To do this, we can make a spectral cube of our calibrated data and compare to the original catalog used for calibration. However, making a spectral cube with fine frequency resolution means that the point-spread function (PSF) side and grating lobes increasingly become a problem due to the sparse sampling of the *uv* plane. Figure 7 shows the HERA Phase I PSF across a wide 60 MHz band (first panel) and three narrower 5 MHz bands (next three panels). For wide-band imaging, the PSF grating lobes are smeared out due to the large bandwidth. However, for narrowband imaging, the grating lobes rise to above 50% of the peak PSF response at image center; for narrower spectral windows, this is only exacerbated. Such strong grating lobes make performing deconvolution to high dynamic range difficult, especially in a confusion-limited regime.

We can partially work around this by applying CLEAN masks around bright sources and then CLEANing down iteratively while opening up the mask to dimmer and dimmer sources. Indeed, this is what we do to make a coarse-channel spectral cube, which consists of MFS images with 5 MHz in bandwidth using iterative CLEAN runs targeting successively dimmer sources. However, in the case of single-channel imaging, even this does not work: the grating lobes are just too severe to deconvolve them from the image without misplacing source flux in unmodeled side lobes. Figure 8 shows the result of a coarse, 5 MHz wide spectral cube CLEAN for a spectral window centered at 155 MHz (gray scale; left). We also show all GLEAM sources in the original model with fluxes above 0.5 Jy in purple, which demonstrates the high degree of confusion given our modest angular resolution: each "source" in our images is generally two or more GLEAM sources blended together. We therefore cannot easily relate the source flux in our images to one or even multiple sources in the GLEAM catalog, as each GLEAM source will have a different contribution to a HERA source given its distance from it and the HERA PSF. If our goal is to compare extracted fluxes between the data and a point-source model, we should take the PSF out of the equation. The deconvolution on the data attempts to do this at some level but is fundamentally limited in precision by the width of the synthesized beam. Another way is to simply add the PSF into the model by imaging the model visibilities and then CLEANing and running a source extraction in the same way as is done for the data. This means that the inherent shortcomings of the deconvolution and limitations of the PSF (both things not really relevant for validating gain calibration done in the *uv* domain rather than the image domain) are kept constant between data and model, so we can make a better comparison between the two.

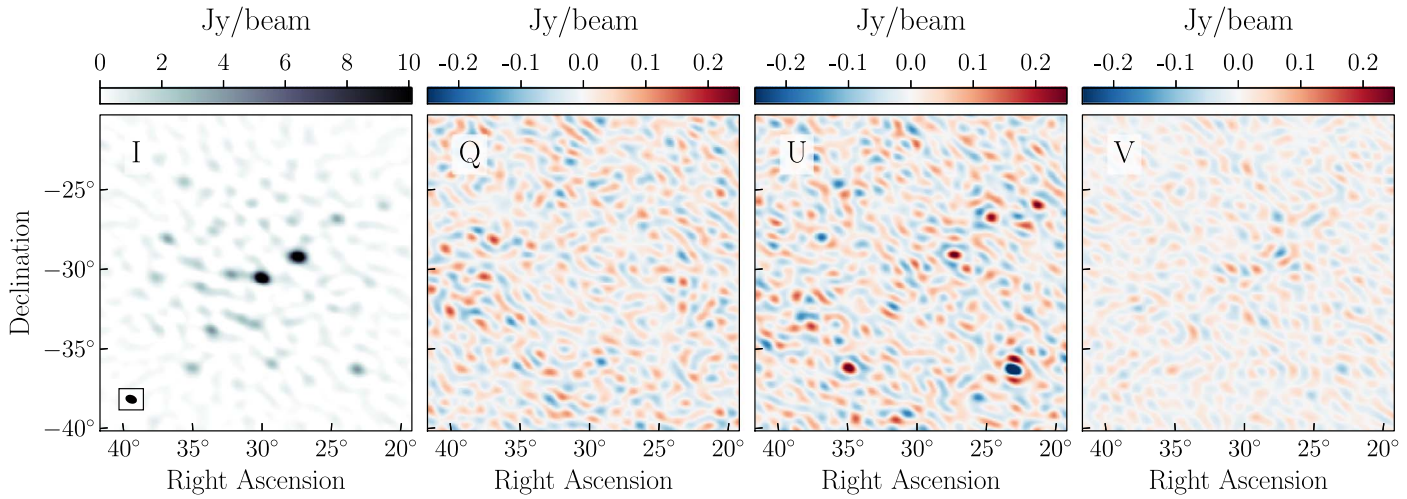

Figure 6. The MFS images (120–180 MHz) of the GLEAM-02H field in all pseudo-Stokes *I* (first panel), *Q* (second panel), *U* (third panel), and *V* (fourth panel) polarizations. Each image is CLEANed with the same parameters down to 1 Jy, with the CLEAN beam shown in the lower left corner. Even with no polarization calibration, the observed leakage from $I \rightarrow Q$, U, and V is a few percent.

Source extraction is done on a source-by-source basis with custom software. First, we select the coordinates of a desired source in the data, then the extraction process makes a postage-stamp cutout in the shape of the synthesized beam with twice its FWHM around the desired source and fits a 2D Gaussian of variable major-axis length, eccentricity, amplitude, and position angle using the `astropy.modeling` module. It then records the integrated flux of the fit in Jy and computes the fit error by taking the rms of the image in an annulus outside the cutout and dividing by the square root of the synthesized beam area (Condon 1997).

This is done for the GLEAM-02H field primary calibrator J0200–3053 for each 5 MHz wide channel in the coarse spectral cube of the data and model, shown in Figure 8. The data (blue) and model (red) are in good agreement with each other across the entirety of the band and relatively good agreement with the primary calibrator's original power-law model from the GLEAM catalog (gray; Table 2). Both the data and model exhibit sinusoidal frequency fluctuations about the power-law model; however, because this structure is represented in model spectra, we can conclude that some of these fluctuations are due to imperfect PSF side-lobe removal in the CLEAN process, rather than calibration errors. If we take the difference between the extracted data and model fluxes, then we see residual deviations at ∼5% of the source's intrinsic flux. However, these deviations look similar in form to the first-order sinusoidal variations about the smooth power law, possibly suggesting that some of these features in the residual are also due to imperfect PSF side-lobe removal. One possibility is that the CLEAN deconvolution achieved better side-lobe removal on the model cube compared to the data cube, which would generate the kind of observed sinusoidal variations in the data-to-model residual. This would not be entirely surprising given the extra terms in the data that are not in the model, including diffuse foregrounds, which would make deconvolution more difficult. The second-order fluctuations in the data-to-model residual (channel-to-channel) hover at roughly 1% of the intrinsic source flux. Overall, these lines of evidence suggest that the quality of the spectral calibration across the band is on the order of a few percent.

However, the leading uncertainty in our absolute calibration is the determination of the overall flux scale. By adopting the GLEAM point-source catalog as our model, we have set the flux scale of our calibration to GLEAM, which itself ties its flux scale to the VLA Low-Frequency Sky Survey redux (VLSSr; Lane et al. 2014), the NRAO VLA Sky Survey (NVSS; Condon et al. 1998), and the Molongo Reference Catalogue (MRC; Large et al. 1981). When comparing their measured source fluxes to sources from these catalogs, their flux scaling appears to be unbiased with an uncertainty of ∼10%. One concern about our usage of a single GLEAM field to set the flux scale is the fact that GLEAM's J0200–3053 source may be an outlier in that distribution, implying that our flux scale could be significantly biased. This concern is tempered by the residual image of Figure 5, which shows that not only J0200–3053 but all sources in the main lobe of the beam have an unbiased residual, meaning that our final flux scale agrees with the GLEAM flux scale for all sources in the main lobe of our primary beam.

To better understand the match between the data and the flux density model, we take the full gain solutions from the GLEAM-02H field and use them to calibrate all baselines in the data. We then form pseudo-Stokes *I* visibilities and coherently average all baselines within a redundant group (i.e., with the same baseline length and orientation). Then we take the Fourier transform of the visibilities across a wide bandwidth spanning 120–180 MHz, having first applied a Blackman–Harris windowing function (Blackman & Tukey 1958) to limit spectral leakage in the discrete Fourier transform (DFT). Before we do this, however, we must first account for the frequency channels that have been flagged due to RFI. These will create strong side lobes in the Fourier transform if not accounted for. To overcome this, we employ a 1D deconvolution algorithm that deconvolves the side lobes due to the RFI, which is conceptually identical to the CLEAN algorithm employed by radio interferometric imaging to interpolate over missing *uv* samples (Högbom 1974) and can be found in the `hera_cal` package. In our case, we build a CLEAN model in delay space out to $\tau = 2000$ ns and interpolate over the flagged channels with the CLEAN model before taking the final DFT to the delay domain, which we do for the data and model

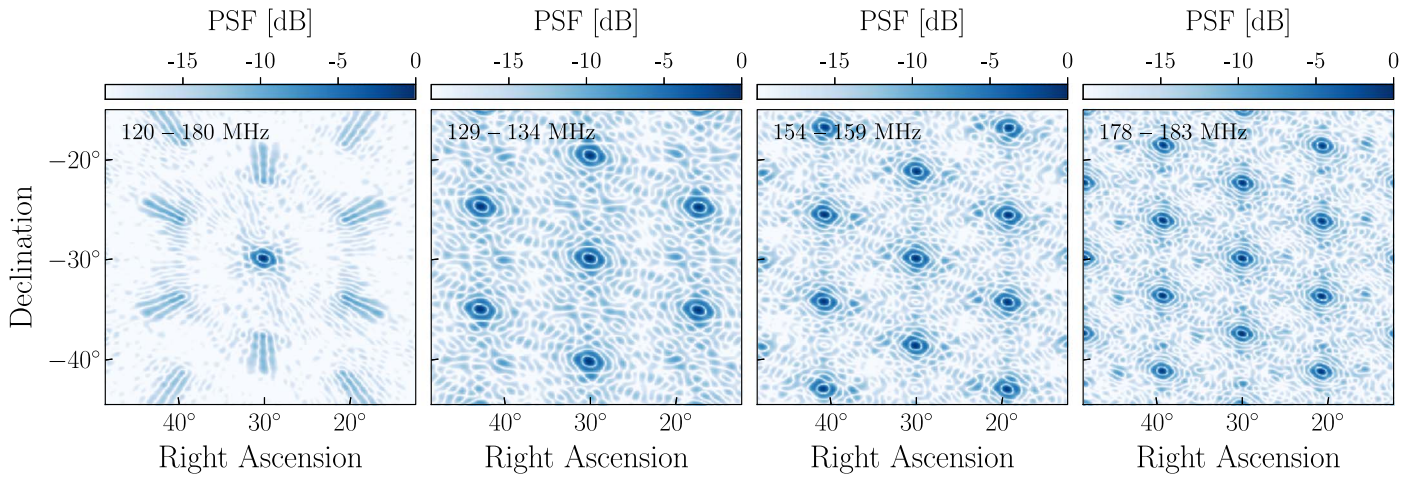

Figure 7. HERA Phase I PSF (without primary beam correction) from a 5 minute observation across a wide band (first panel) and a narrow band located in a low-band (second panel), mid-band (third panel) and high-band (fourth panel) spectral window. The grating lobes of the narrowband spectral windows appear in hexagonal patterns reflecting the (un)sampled *uv* spacings on the array and reach upward of 50% of the peak PSF response at image center.

visibilities in an identical fashion. We then coherently time-average the 5 minutes of data around the GLEAM-02H calibration field, take the absolute value of the averaged visibilities, and average all baselines of the same length, regardless of orientation. This is the same procedure one would use to form 2D cylindrically averaged power spectra, but in this case, we are working with just the visibilities in the Fourier domain.

Figure 9 shows this for the calibrated data (left) and model (middle) and their residual (right). From it, we can clearly see the pitchfork-like foreground wedge with a main component centered at $\tau = 0$ ns and branches at positive and negative delay following the horizon line of the array, which is not plotted for visual clarity and is explained in more detail in Section 6. Recall that baselines shorter than 40 m in length (white dashed lines) are not used in calibration. The pitchfork branches are caused by the foreshortening of a baseline's separation vector at the horizon, thus increasing its sensitivity to diffuse emission (Thyagarajan et al. 2015). The point-source model, lacking diffuse foregrounds, clearly does not have a strong pitchfork feature. This discrepancy will create gain errors in the calibration solutions at the delay scale of the pitchfork, which, for baselines above 40 m, begins at around 150 ns and extends beyond that for longer baselines. This is explored in the following section. Lastly, the data and model are somewhat well matched at $\tau \sim 0$ ns, with the residual power being suppressed by a factor of 10 compared to the data but still above the noise floor of the data outside the wedge. This residual power can come from unmodeled diffuse flux in the main lobe of the primary beam but is also likely to be from calibration errors due to mismodeled point sources. An increase in observed power in the data at large delays, $|\tau| > 800$ ns, is a cross-coupling systematic, not foreground signal (Kern et al. 2019a).

4. Gain Stability

In this section, we characterize the spectral and temporal properties of the derived complex antenna gains and discuss their impact on downstream analyses. Gain calibration is a multiplicative term in the frequency and time domain, meaning it can equivalently be thought of as convolution in the Fourier domains of delay and fringe rate; the Fourier duals of frequency and time, respectively, by a "gain kernel"; or the Fourier transform of the gain response. Solving for and applying antenna-based gains can therefore be thought of as trying to deconvolve the inherent gain kernel imparted by the instrument. For 21 cm experiments aiming to uncover a signal buried under noise and systematics, the principal concern when applying gain solutions to the data is understanding how this gain kernel may or may not be smearing foreground signal to spectral modes that are otherwise foreground-free; any kind of deviation in the derived gain solution from the true underlying gain will cause such smearing at some level.

4.1. Spectral Response

Works investigating sky-based calibration in the limit of an incomplete sky model showed that it results in gains with erroneous spectral structure that can fundamentally limit 21 cm studies (Barry et al. 2016; Ewall-Wice et al. 2017; Byrne et al. 2019). Similar effects have been shown to exist for redundant calibration, where inherent nonredundancies of the array create a similar type of spectrally dependent gain error (Orosz et al. 2019). What has yet to be studied in detail is how other kinds of instrumental systematics, such as mutual coupling or cross talk, get picked up in the process of gain calibration and what their effect is in shaping the inherent and estimated gain kernel. For systematics like cross talk and mutual coupling, which are highly baseline-dependent, one would naively expect that the antenna-based gains would not significantly pick up on these terms due to their decoherence when averaged across different baselines; however, it would not be surprising to see them reflected in the gain solutions at some level, even if they are averaged down to some degree. Furthermore, Figure 9 shows us that there is a nonnegligible data-to-model discrepancy caused by unmodeled diffuse emission even for baselines above our 40 m cut, which will also create gain errors.

To summarize Kern et al. (2019a), the HERA Phase I system shows evidence for cross-coupling systematics at large delays, $|\tau| > 800$ ns, as well as for diffuse flux and/or mutual coupling at smaller delays corresponding to a baseline's geometric horizon (for $|\boldsymbol{b}| = 45$ m, this is $\sim$150 ns). In Figure 10, we show the frequency and delay response of the CASA-derived, sky-based gains from Section 3.2. We plot the gain amplitude (top left), the gain phase after removing the cable delay for

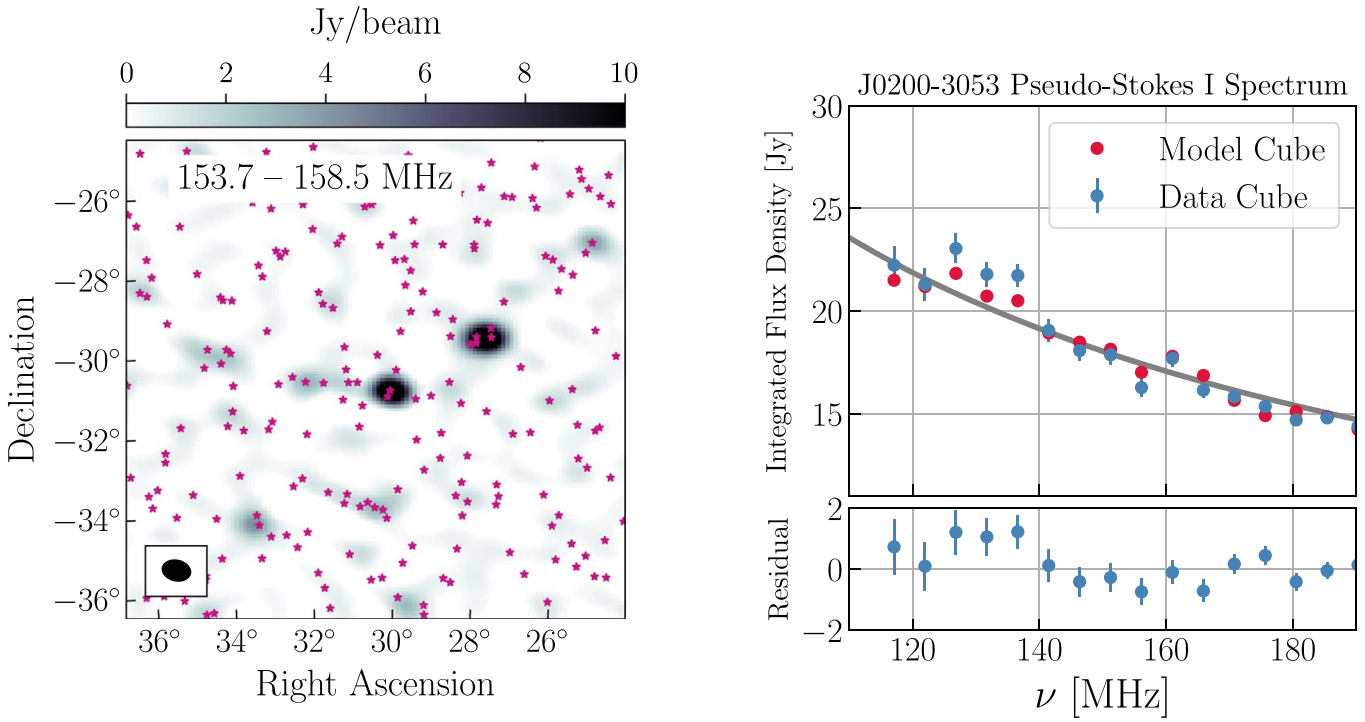

Figure 8. Extracted spectrum of the primary calibrator GLEAM J0200–3053 from the GLEAM-02H field. Left: CLEANed MFS image of the data (color scale) across a narrow band (153.7–158.5 MHz). The purple markers show each GLEAM point source above 0.5 Jy used in the initial model, demonstrating the degree of source confusion given the Phase I angular resolution. Right: extracted spectrum of J0200–3053 (center of left image) across each channel in the data (blue) and model (red) spectral cubes. The data and model are in good agreement with each other and well fit by the original input GLEAM J0200–3053 power-law model (gray). Large-scale frequency deviations from the power-law fit are partially reflected in both the data and model, suggesting that they are not due to miscalibration but to imperfect PSF side-lobe removal in the process of imaging. The data cube—model cube difference shows residual structure at the ∼5% level.

each antenna (top right), and the Fourier transform of the gains in delay space (bottom) normalized to their peak power at $\tau = 0$ ns. We categorize the gains into Type 1 (blue) and Type 2 (red) signal chains (Figure 2), which shows a clear bimodality in the spectral structure of the gains between these groups. This bimodality is also seen in the reflection properties of the signal chains and discussed in more detail in Kern et al. (2019a). Arrows mark the expected regions in delay space where certain electromagnetic elements in the signal chain can create systematics, such as reflections within the 14 m dish and in the 20 and 150 m coaxial cables. The gain kernels of each antenna (Figure 10, bottom) also clearly show that instrumental cross-coupling systematics at $|\tau| > 800$ ns are being picked up by the gain solutions. It also shows unmodeled diffuse emission at lower delays, $|\tau| > 200$ ns, which may also have contributions from mutual coupling systematics that appear at similar delays. Because instrumental cross-coupling and diffuse emission are baseline-based and not antenna-based, they cannot be calibrated out of the data with antenna-based, direction-independent gains and must be removed on a per-baseline basis. This means that the presence of these structures in the gains will only spread the systematics around: in the worst case, spreading them to baselines that may have been systematic-free to begin with.

Figure 11 shows the result of applying sky-based gains to the visibility data and transforming to the Fourier domains of delay and fringe-rate space. We apply the gains to 8 hr of drift-scan data from a single 29 m east–west visibility, and we do this having filtered the gains in three different ways: (1) the first method (simple calibration) takes only the band-averaged amplitude and cable delay component of the gain, (2) the second method (full calibration) takes just the full gain solution as is, and (3) the third method (smooth calibration) smooths the gains across the frequency out to a 100 ns scale, which is also plotted in Figure 10 (black dashed line). The bottom panel shows the time-averaged delay response of the panels shown above. In the simple calibrated data, the foregrounds are contained to low delays and appear predominately at positive fringe rates, which we expect because the sky rotates in a single coherent direction in the main lobe of the primary beam (Parsons et al. 2016). Foregrounds can also occupy near-zero and negative fringe-rate modes, which correspond to structures on the sky near the south celestial pole and the horizon but are attenuated by the primary beam response. If the data were nominal, then the rest of the Fourier space would be dominated by thermal noise; however, this is not what we observe. We also see cable reflection signatures, which should appear as reflected copies of the foregrounds at the same fringe rates but at positive and negative delays (marked). And we see strong cross-coupling features at large positive and negative delays occupying near-zero fringe-rate modes (marked).

When we go to apply the full calibration, we find a large amount of excess structure at intermediate and large delays occupying positive fringe rates, which is not surprising given the gain kernels shown in Figure 10. We see that other baselines that happened to have the systematics at intermediate delays have contaminated this baseline at the same delays. What is more, these systematics are now occupying the same

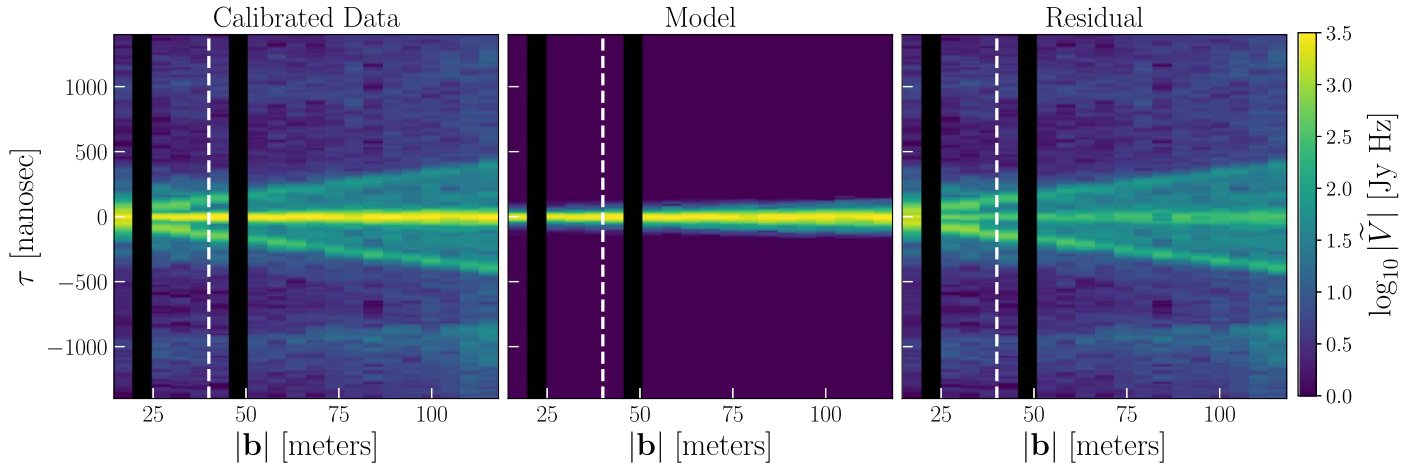

Figure 9. Redundantly averaged pseudo-Stokes *I* visibilities in delay space transformed across a 120–180 MHz spectral window, ordered according to baseline length. We show the calibrated data (left) and point-source calibration model (middle) and their residual (right). Short baselines to the left of the white dashed line are not used in calibration. Black regions represent a lack of data at those baseline lengths. The data clearly show a pitchfork-like foreground wedge predicted by Thyagarajan et al. (2016). Note that the edges of the pitchfork are not reflected in the calibration model, which will generate calibration errors. The residual power of the main foreground lobe in the wedge is suppressed by about a factor of 10 compared to the data but is still seen above the noise floor of the data. Additional power at large delays ($\tau \sim 1000$ ns) is the same systematics seen in Kern et al. (2019a).

positive fringe-rate modes as the sky[26] and therefore cannot be easily removed with standard cross-coupling removal techniques (Kern et al. 2019b). There are some benefits of the full calibration, though. One is that it can calibrate out signal chain reflections because those factor as antenna-based terms. This can be seen in the data as the suppression of the cable reflections at large positive and negative delays, as well as in the suppression of the dish reflection at $\tau = \pm 50$ ns (which is most apparent as the tightening up of the contours in the brightest spots of the foregrounds or the drop in the shoulder power in the time-averaged spectra). While cable reflections at high delays can be calibrated out with sky-agnostic modeling (Ewall-Wice et al. 2016a; Kern et al. 2019b), calibrating out reflections at low delays that bleed into the main foreground lobe is harder and thus better suited to correction via standard gain calibration.

The ideal compromise, then, is to smooth our gains to keep the gain kernel at low delays and suppress its power at delays that we no longer trust its response. For the calibration at hand, this seems to be at roughly 100 ns, which enables the calibration to pick up on the dish reflection at 50 ns but suppresses the spurious terms in the gains at 150 ns and beyond. Given our 100 MHz bandwidth with 1024 channelization, a maximum delay range of 100 ns leads to about 15 free delay modes in the smoothed gains, which can be thought of as a smoothed gain with 15 spectral degrees of freedom per antenna and dipole polarization. Applying this gain to the data (last two panels in Figure 11), we see that we recover the best of both scenarios: the dish reflection is suppressed as desired, and we also do not spread more instrument coupling at intermediate and high delays over what is already present in the data. To perform the smoothing, we use the same delay domain deconvolution technique described before as a low-pass Fourier filter, which is useful given that the gains are also flagged at certain frequency channels due to RFI. Although this calibration is performed for a single time, one can also take time- and frequency-dependent calibration solutions and smooth across both the temporal and spectral axes with this technique.

We can also show the effects of the smooth and full calibration on the full data set. We do this by applying the calibration to the data and transforming them to the delay domain in a similar manner as was done for Figure 9. In this case, Figure 12 plots this for the smooth (left) and full (middle) calibrated data and their fractional residual (right). Note that the calibrated data are plotted on the same color scale as Figure 9 but with a smaller delay range to highlight features within the foreground wedge. We see that the two calibrations achieve a good match at low delays, as expected, but for delays beyond the smoothing scale, we find that the full calibrated data have significant excess structure (red) compared to the smooth calibrated data. This is indicative of the full calibration introducing spectral features into the data, rather than calibrating them out, which is highly suggestive of gain errors on these scales and further motivates the $\tau \sim 100$ ns smoothing scale of the gains derived in Section 3.2.

Philosophically, this kind of approach to gain calibration—in other words, keeping only degrees of freedom like low delay modes that we trust and filtering out the rest—is conservative from the perspective of not introducing structure into the data that was not already there. The cost of this approach is that we are not calibrating out gain structure at these delays inherently introduced by the instrument, if it exists in the first place. At the moment, however, we do not really have much of a choice: providing a constrained calibration with a few degrees of freedom is the best we can currently do, and until we have evidence that structure in the gain kernel at higher delays is real gain structure, we should not attempt to calibrate it out. This approach makes interpreting a fiducial detection in the power spectrum at similar intermediate delays somewhat convoluted, and a suite of null tests and jackknives will be necessary to try to tease out whether said detection is residual calibration structure or real sky structure.

The obvious question moving forward for HERA, then, is: do we believe that there is true gain structure at low and intermediate delays that we need to calibrate out? The answer to this depends on the required dynamic range. For low delays,

[26] This happens because the gains are a multiplicative term, meaning that although the systematics originally occupied $f \sim 0$ Hz, the contaminated gains spread them to $f > 0$ Hz modes.

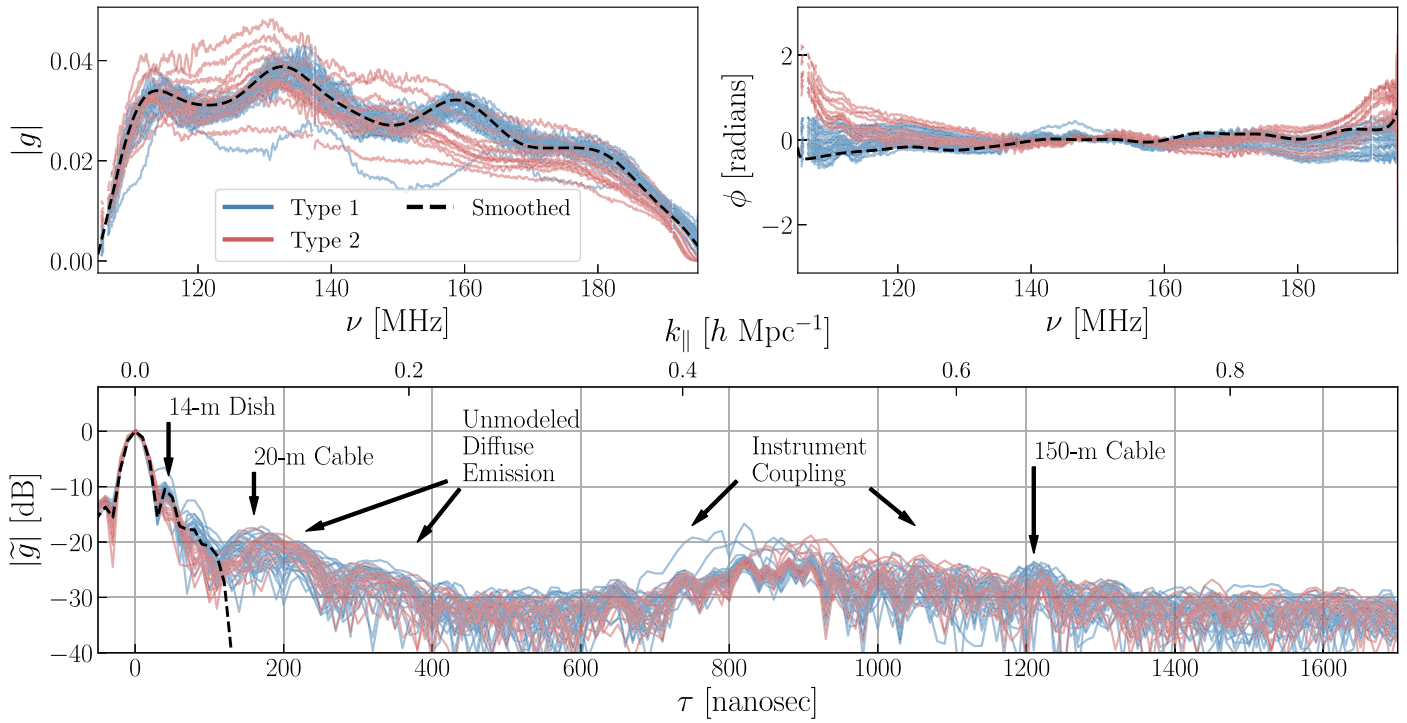

Figure 10. Antenna gains derived from the GLEAM-02H field. Type 1 and 2 signal chains are plotted in blue and red, respectively. The phases of the gains (top right) are plotted after taking out the cable delay from each antenna for visual clarity. The peak-normalized delay responses of the gains show structure at delays representative of elements in the signal chain (bottom), as well as contamination by terms that are not antenna-based, like unmodeled diffuse emission and instrumental coupling systematics (Kern et al. 2019a). We also show one of the Type 1 gains smoothed at a 100 ns scale for reference (dashed black line).

we generally need 10^5 in dynamic range performance of the gain kernel due to the foreground-to-EoR amplitude ratio; for larger delays, this requirement becomes more stringent as the EoR signal is expected to weaken. Therefore, do we think HERA has true gain structure at some level above -50 dB at $\tau \sim 200$ ns? Based on simulations (Fagnoni et al. 2019) and a rough extrapolation of Figure 10, the answer is probably yes; therefore, we need a way to remove the cross-coupling systematics from the data before performing antenna gain calibration. Cross-coupling systematic removal is done by applying a high-pass filter in fringe-rate space (Kern et al. 2019b; Kolopanis et al. 2019). This removes cross-coupling, which occupies low fringe rates, but it also removes a component of the foregrounds as well, which we need for calibration. Doing this only on the data and not on the model would create a discrepancy in the data that would act as its own form of systematic. Fringe-rate filtering therefore needs to be done on the model and data before calibration in order to probe the true instrument gain kernel to higher and higher delays. Achieving high fringe-rate resolution for a high-pass filter means simulating a large LST coverage with a wide-field flux density model. Unfortunately, the CASA-based calibration methodology presented in this work does not easily lend itself to this, as it only reliably simulates short time intervals near the calibration field. This kind of analysis is best done using a numerical visibility simulator with wide-field diffuse and point-source maps, which we defer to future work.

Other smoothing algorithms have been investigated in the literature, which has been motivated due to a recent understanding of how incomplete sky models cause gain errors in sky-based calibration (Barry et al. 2016; Byrne et al. 2019) and nonredundancies cause gain errors in redundant calibration (Ewall-Wice et al. 2017; Orosz et al. 2019). The MWA, for example, uses low-order polynomials to smooth their sky-based gain solutions to limit gain error spectral structure in 21 cm power spectral analyses (Beardsley et al. 2016; Barry et al. 2019a). The reason we opt for direct Fourier filtering of the gains in this work is because a truncated polynomial basis is not able to encapsulate arbitrary gain fluctuations on large scales; in other words, they do not form a complete basis in the Fourier domain for low delay modes. This is fine for mitigating small-scale structure but means one runs the risk of not calibrating out large-scale modes that can cause biases in narrowband power spectrum analyses, although in simulated MWA analyses, there is no evidence for such biases (Barry et al. 2016).

4.2. Temporal Response

In this section, we use the data to assess the temporal stability of the instrumental gain. HERA observations are taken in drift-scan mode, meaning the array does not change or move over the course of observations. This lends itself to a fairly stable instrument as a function of time, and we therefore do not expect large deviations in the gains over short time intervals. However, effects such as ambient temperature drift and the cooling cycle within signal chain nodes are known to cause slight drifts in the calibration over the course of a night (e.g., Jacobs et al. 2013). In this section, we investigate the data to quantify the amplitude of these gain drift terms and confirm they can be calibrated out if necessary. Note that we do not actually apply time-dependent gains to the data in the remainder of this work; we merely present ways in which these terms can be calibrated out for deep integrations if necessary.

All signal chains in the HERA array are brought via coaxial cable to an RFI-shielded and air-conditioned container in the

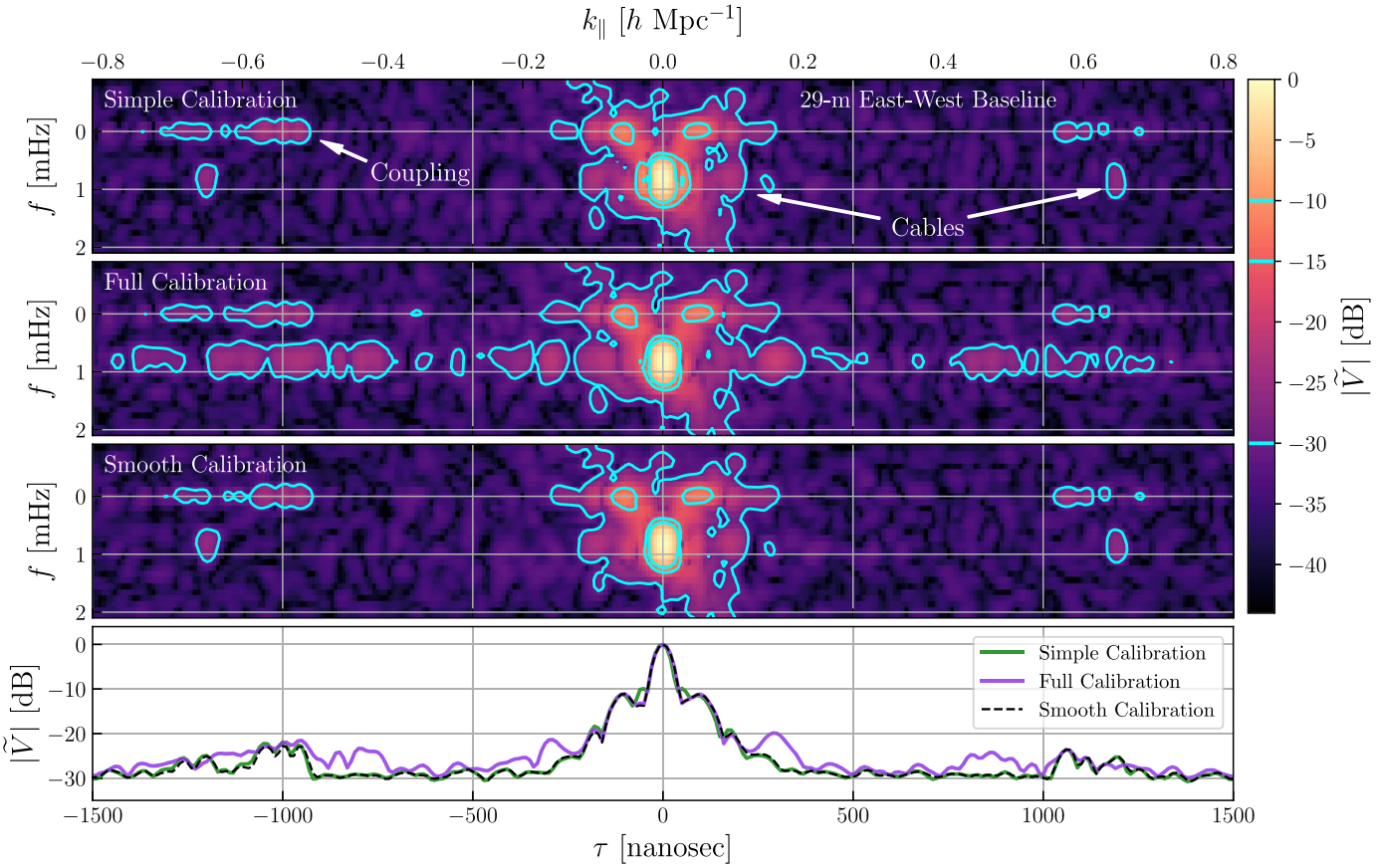

Figure 11. Sky-based gains applied to a single 29 m east–west visibility over 8 hr of LST and transformed to delay and fringe-rate space (see text for details). The data are peak-normalized, and the contours show −30, −15, and −10 dB levels. The time-averaged delay responses are shown in the bottom panel. Sources of cable reflection and instrument coupling are marked. The full gain applied to the data leads to significant contamination of coupling systematics due to the full gain kernel smearing the foreground horizontally in fringe-rate and delay space. Smoothing the calibration allows us to calibrate out the features at low delays that we know can be calibrated (e.g., dish reflections) and toss out features in the gain kernel above 100 ns.

field, where the data are converted from analog to digital signals and then correlated. Due to the air-conditioning cycle within this container, which cycles at roughly a 6 minute period, we expect the overall amplitude of the gains to drift at the same timescales. We can estimate the amplitude of this drift using a smoothed version of the autocorrelations; this is the approach adopted by the LWA (Eastwood et al. 2019), which faces the same issue. Assuming that the only temporal structure in the autocorrelations occurs intrinsically at the timescale of the beam crossing time (∼40 minutes), we can probe the time structure from the gains by taking a time-smoothed version of the autocorrelation and dividing it by the unsmoothed autocorrelation. We smooth a handful of autocorrelation visibilities on a 20 minute timescale, divide their unsmoothed visibility counterparts by them, and take their square root, which leaves us with a set of ratio waterfalls as a function of frequency and time for each antenna polarization. We show some of these in Figure 13, which plots the square root ratio for each time and frequency bin for four antennas (top panels) and their frequency average as a function of LST (bottom panel). We see that the gain fluctuation induced by the air-conditioning cycle in the container has a coherent phase and amplitude across all antennas and polarizations and is also fairly constant across frequency. The frequency average of each antenna and its respective average is shown to reflect a sawtooth profile as a function of time, whose profile inversely matches temperature data collected within the container. Figure 13 shows us that the 6 minute gain oscillations are a very small effect at the 0.1% level and can be decently well calibrated by a single number as a function of time for all antennas, polarizations, and frequency channels in the array. The HERA Phase II configuration will have a forced air cycling system that will better control fast temperature variations in container units.

A steady decrease in ambient temperature after sunset can cause slow evolution in the performance of the exposed part of the signal chains, in particular the low-noise amplifier in the FEM, which is attached to the feed. This kind of gain drift is expected to be slow but could add up over the course of an entire night of observing, especially if we choose to calibrate the data once at either the beginning or end of the night. To test this, we calibrate a single night of data at three different fields (Figure 3) at different times of a single night and compare the average gain amplitude derived from each field. Figure 14 shows this drift having normalized the gains to the 2 hr field, demonstrating a slow drift that over the course of ∼5 hr leads to about a 10% drift in the gain amplitude. Also plotted is the ambient temperature measured by a nearby weather station, which shows an expected inverse correlation with the antenna gain. Similarly, the band-averaged gain phase drift (after taking out the cable delay) is kept to within 0.2 rad over the same time interval, but unlike the average amplitude, the phase drift does not appear monotonic in time.

Using the temperature data, we can derive an ambient temperature coefficient for the change in the average gain as a

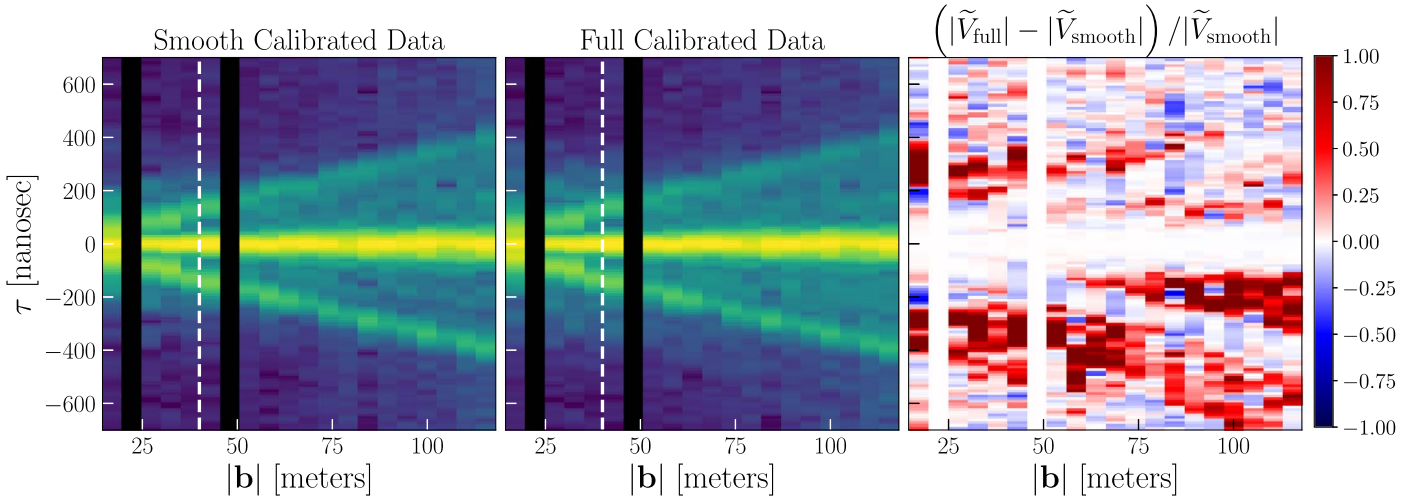

Figure 12. Redundantly averaged pseudo-Stokes I visibilities in delay space transformed across a 120–180 MHz spectral window, ordered according to baseline length. We show the smooth (left) and full (middle) calibrated data and their fractional residual. The calibrated data are plotted on the same color scale as Figure 9 over a smaller delay range to highlight the features within the foreground wedge. Within the smoothing scale of 100 ns, the fractional residual shows that the two are in good agreement, as expected. Outside the smoothing scale, however, the residual shows significant excess structure (red) in the full calibrated data that is not seen in the smooth calibrated data, which suggests that the structures are not real and are errors in the gain solution.

function of temperature difference (Jacobs et al. 2013). We can represent a relationship for the difference in ambient temperature relative to the ratio of the derived gain response of the analog system as

$$10 \cdot \log_{10} \left| \frac{g_{\rm new}}{g_{\rm norm}} \right| [\rm dB] = C \cdot (T_{\rm new} - T_{\rm norm}) \, [\rm K], \qquad (3)$$

where $T_{\rm norm}$ and $g_{\rm norm}$ are the ambient temperature and average gain amplitude at the time of gain calibration (i.e., our normalization time), $T_{\rm new}$ and $g_{\rm new}$ are the temperature and gain at any new time in LST, and C is the temperature coefficient in units dB $\rm K^{-1}$. In Figure 14, for example, we have chosen the normalization to be at 2 hr LST. Using the three data points from Figure 14, we derive a temperature coefficient of -0.031 dB $\rm K^{-1}$ for the gains. With a similar approach, Pober et al. (2012) also derived a gain temperature coefficient of -0.03 dB $\rm K^{-1}$ for the PAPER system, which used similar front-end hardware as HERA Phase I. Jacobs et al. (2013) also used data from two different seasons to derive an autocorrelation temperature coefficient for the PAPER system of -0.06 dB $\rm K^{-1}$, which, when divided by a factor of 2 in order to map it to a gain temperature coefficient, is also in agreement with these results.

5. Combining Redundant Calibration

A key part of HERA's design is to exploit its inherent redundant sampling of the *uv* plane for precision redundant calibration (Dillon & Parsons 2016). Redundant calibration asserts that all visibilities of the same baseline length and orientation (uniquely defining a "baseline type") measure the same visibility, which with enough redundant baselines allows for an overconstrained system of equation while keeping the true visibility a free parameter (Wieringa 1992; Liu et al. 2010). This means that redundant calibration does not need an estimate of the true model visibilities and thus temporarily skirts some of the issues with incomplete or inaccurate sky models. In practice, this is never exactly true, and slight antenna position and primary beam uncertainties therefore generate gain errors in redundant calibration (Ewall-Wice et al. 2017; Orosz et al. 2019). Nonetheless, we would like to explore options for combining redundant and absolute calibration to exploit their complementary advantages, either as an alternative or a hybrid calibration pipeline.

For a baseline between antennas i and j and another between antennas j and k, both belonging to the same baseline type of ij (for example, antenna pairs (23, 24) and (24, 25) from Figure 2), the redundant calibration equations are

$$\begin{aligned} V_{ij}^{\rm data} &= g_i V_{ij}^{\rm model} g_j^* + n_{ij} \\ V_{jk}^{\rm data} &= g_j V_{ij}^{\rm model} g_k^* + n_{jk} \\ &\vdots \end{aligned} \qquad (4)$$

Note that the model visibility for V_{jk} is now $V_{ij}^{\rm model}$. In this case, we are left with four free parameters, g_i, g_j, g_k, and $V_{ij}^{\rm model}$, which we can solve for by minimizing their χ^2,

$$\chi^2 = \sum_{i,j} \frac{|V_{ij}^{\rm data} - g_i V_{ij}^{\rm model} g_j^*|}{\sigma_{ij}^2}, \qquad (5)$$

where σ_{ij}^2 is the noise variance on baseline ij and the sum is over all antenna pairs in the array. Although a two-baseline array like the one in Equation (4) cannot be redundantly calibrated, we can see that increasing the number of redundant baselines will turn this into an overconstrained system of equations (Liu et al. 2010). However, redundant calibration is not the final answer for antenna-based calibration, as there exist fundamental degeneracies that redundant calibration simply cannot constrain. One of these degeneracies is the average gain and model visibility amplitude. Looking at Equation (5), we can see that if we multiply all antenna gains by some fraction A and then divide all model visibilities by A^2, we leave the final χ^2 unchanged. Recall that we are free to do this because, unlike in sky-based calibration, the model visibility is a free parameter. Thus, it can perfectly counteract such deviations in the gains and implies that the full system of equations is

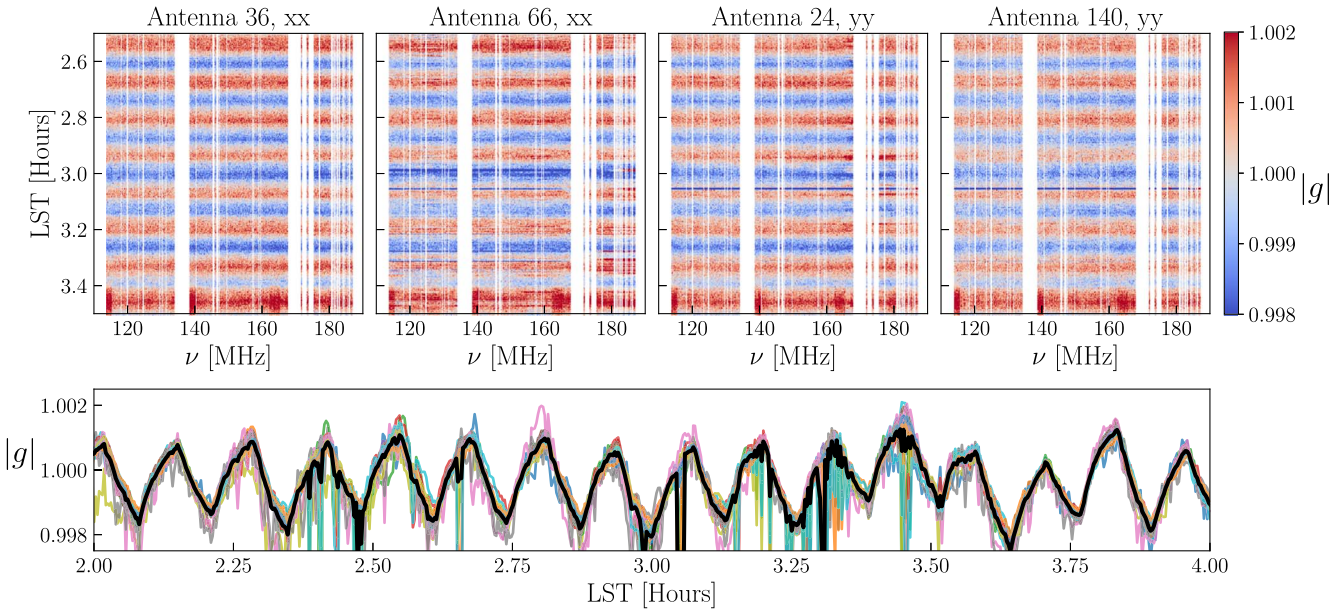

Figure 13. Temperature oscillations in the instrumental gain due to an air-conditioning cycle in the field container housing the ADC are a 0.1% effect. The top panels show the square root of the ratio of the raw autocorrelations to the time-smoothed autocorrelations for a few antennas and both *XX* and *YY* polarization. The oscillation looks to be of roughly the same amplitude across different antennas, polarizations, and frequencies. The bottom panel shows the frequency-averaged oscillation for a handful of antennas (colored lines) and their average (black). This shows a sawtooth time profile that also matches temperature data collected in the container.

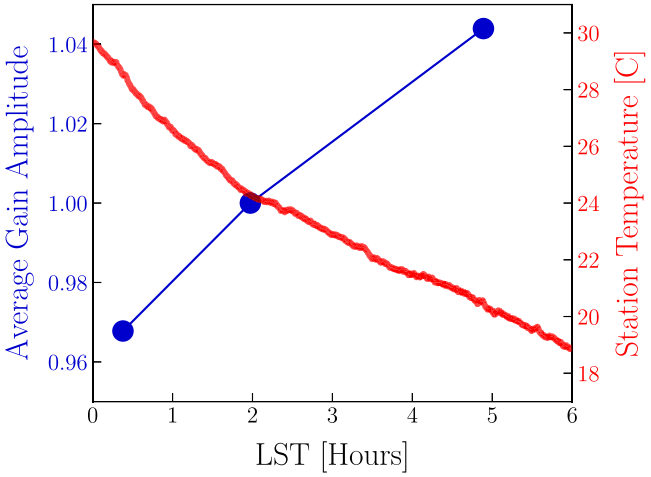

Figure 14. Average gain amplitude drift (blue) throughout the 2458098 observing night, derived from three independent calibration fields and normalized to the field at 2 hr. We also overplot the ambient temperature measured by a nearby weather station (red), showing an expected inverse correlation with the gain drift. Using Equation (3), these data yield a gain temperature coefficient of -0.031 dB K^{-1}.

insensitive to their average amplitude. In addition to the average gain amplitude, the other major degeneracy associated with redundant calibration is known as the "tip-tilt" phase gradient across the east–west and north–south coordinates of the array (Zheng et al. 2014; Dillon et al. 2018). If each antenna is assigned a vector $\boldsymbol{r}_i$ originating from the center of the array to its topocentric coordinates of east and north, we can insert a "tip-tilt" phase gradient into the gains as

$$g_i \rightarrow g_i \exp(i\boldsymbol{\Phi}\boldsymbol{r}_i)$$
$$\text{where } \boldsymbol{\Phi} = (\Phi_E, \Phi_N). \tag{6}$$

The coefficient $\boldsymbol{\Phi}$ is therefore a phase gradient coefficient with units of radians per meter, with separate coefficients for the east and north directions. Such a perturbation to the gains is a degeneracy in redundant calibration because we can exactly cancel this out by applying the opposite term to the model visibilities. For example, we can express the second term in the χ^2 metric of Equation (5) as

$$\begin{aligned} & g_i V_{ij}^{\text{model}} g_j^* \rightarrow \\ & g_i \exp(i\boldsymbol{\Phi}\boldsymbol{r}_i) V_{ij}^{\text{model}} \exp(-i\boldsymbol{\Phi}\boldsymbol{r}_{ij}) g_j^* \exp(-i\boldsymbol{\Phi}\boldsymbol{r}_j) \\ & = g_i V_{ij}^{\text{model}} g_j^* \exp(i\boldsymbol{\Phi}\boldsymbol{r}_{ij}) \exp(-i\boldsymbol{\Phi}\boldsymbol{r}_{ij}) \\ & = g_i V_{ij}^{\text{model}} g_j^*, \end{aligned} \tag{7}$$

where we have made use of the fact that $\boldsymbol{r}_i - \boldsymbol{r}_j = \boldsymbol{r}_{ij}$ and can see that after substitutions, the term is unchanged. This amounts to a total of three parameters, the average gain amplitude and the east and north phase gradients, that need to be solved for after redundant calibration and require a sky model to pin down (per frequency, time, and polarization). Thus, the issues of inaccurate sky models are somewhat mitigated but not totally circumvented by redundant calibration (Byrne et al. 2019). From a sky-based calibration perspective, these degeneracies can roughly be thought of as the overall flux scale of the array and its pointing center on the sky.

There are multiple ways to fill in the missing degenerate parameters of redundant calibration. One approach is to take the redundant calibration solutions and project only their degenerate components onto the degenerate modes in the sky-based calibration solutions (Li et al. 2018). One can also take model visibilities and set up a new calibration equation that solves explicitly for the degenerate parameters (partial absolute calibration). Finally, one can take the sky-based calibrations as a starting point by applying them to the data and then run redundant calibration. The latter two of these, along with standard sky calibration, are shown in Figure 15, outlining the order of operations of the three proposed calibration schemes:

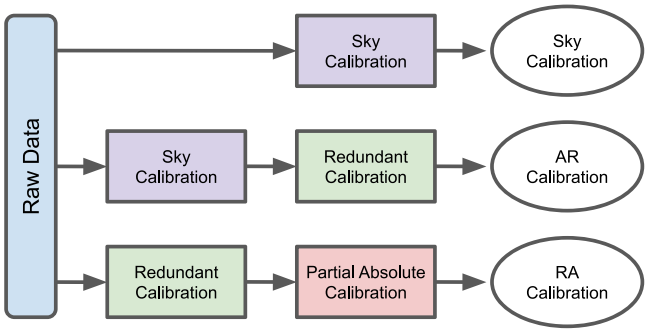

Figure 15. Schematic showing the order of operations for three related calibration strategies, similar to that of Li et al. (2018). For AR and RA calibration, the gains from the first step are applied to the data before proceeding to the second step. In addition, the gains derived by redundant calibration have their degenerate modes projected out before proceeding.

sky calibration, sky + redundant (AR) calibration, and redundant + partial absolute (RA) calibration. Both AR and RA calibration schemes are built into the `redcal` module of the `hera_cal` software package (which we use here), including setting up and solving a system of equations that specifically picks out the degenerate parameters of redundant calibration given a set of sky model visibilities, which we discuss in more detail in Appendix A. For the RA approach discussed here, the model visibilities used for extracting the degenerate modes are simply the raw data calibrated with the sky-based gains.

Because redundant calibration cannot constrain the degenerate modes inherent to its system of equations, the output gains will generally have some random combination of degenerate vectors, which will be influenced by the convergence of the calibration solver and its starting point from the raw data. To fix this, we can project out these degeneracies by fixing them to some a priori chosen position, which will then get filled in by absolute calibration (Dillon et al. 2018; Li et al. 2018). The simplest thing is to rescale the gains such that the average amplitude is 1.0 and the phase gradient is 0.0, which is done to just the redundant calibration portion of the gains in both RA and AR calibration.

We saw in Figure 10 the presence of antenna-based structures that we expect to appear in the gains, like the dish reflection and the 20 and 150 m cable reflections, but we also saw significant contamination by instrumental coupling across a wide range of delays. To understand the kinds of structures picked up by redundant calibration, we can inspect the gains in a similar manner. Figure 16 shows the distribution of the gains at each step in the AR calibration scheme in amplitude (left) and phase (center), having removed the cable delay for each antenna. It also shows the gains Fourier-transformed across frequency in delay space (right) and peak-normalized. The top panels of Figure 16 show just sky calibration (the same as Figure 10). The middle panels show just the redundant calibration component of the gain, where, in deriving them, we first apply the sky calibration gains to the data, and the bottom panels show the final product of the two gains. Note that the redundant calibration gains derived here use the same baselines as the sky calibration of $|\boldsymbol{b}| > 40$ m. For the redundant calibration gains, we can see that the average amplitude is 1, as expected, and has similar kinds of spectral structure as the sky calibration gains. Looking at their product, or the AR calibration gains, we can see some of the benefits of redundant calibration. Compared to the sky calibration delay response (purple), the AR delay response (orange) has a slightly suppressed bump at ∼200 ns, which can also be seen as the negation of the coherent ripple in the middle row of phase plots. We observe this ripple in the sky-based gain phases (top), which seems to be corrected for by redundant calibration (middle) such that their product (bottom) demonstrates less of a ripple. One possible explanation is that this ripple is caused by an imperfect sky model that creates spectral errors in the sky-based gain and is then corrected by redundant calibration. However, we still see significant power at $\tau \gtrsim 200$ ns, which could originate from nonredundancies between nominally redundant baselines specifically at the horizon, where diffuse emission generates the pitchfork effect in the data but also the per-antenna primary beams are likely the least redundant with each other. Similar to how unmodeled diffuse emission created gain errors in sky calibration, these kinds of nonredundancies will create errors in redundant calibration and appear at similar delays (Orosz et al. 2019). Previous work showed that nonredundancy seems to be worse for short baselines (Carilli et al. 2018), but quantifying this in more detail is still in progress (Dillon et al. 2020, in preparation). The AR calibration gains also show significant power at $\tau \gtrsim 800$ ns, which shows that redundant calibration is not immune to picking up cross-coupling instrumental systematics. The RA gain solutions show nearly the same structure as solutions derived from AR calibration down to below 1% in fractional difference, so we do not plot them here for brevity.

To further show the effects of redundant calibration on the data, we take the full gains (this time from the RA calibration scheme) and apply them to the data. We then redundantly average and Fourier transform them, similar to Figure 9. Figure 17 shows this process having applied the full sky and RA calibration, as well as the fractional difference between the two (right). Areas where the RA calibration is introducing new structures show up as red, and areas where it is calibrating out structures compared to the sky calibration show up as blue. We see that for shorter baselines and delays near $|\tau| \sim 200$ ns, RA calibration is inserting less structure into the data compared to sky calibration, which agrees with our observation earlier that spurious gain structure at those delay scales seem suppressed. However, we see that at slightly smaller delays and larger baseline lengths, RA calibration is also inserting additional power compared to sky calibration, which is likely a result of its own gain errors. Additionally, at small delays ($|\tau| < 100$ ns), the two are in good agreement with each other.

The takeaway from this section is that (1) all three calibration schemes yield gains that are similar at low delays, (2) hybrid-redundant calibration seems to correct for some of the errors in the sky-based calibration but still introduces its own set of errors, and (3) both sky and redundant calibration suffer from gain errors that are induced by baseline-dependent instrumental systematics. Moving forward, future analyses will benefit from attempting to model diffuse emission and removing instrumental cross-coupling systematics before calibration in order to calibrate intermediate delay scales and exploit the full power of a combined redundant and absolute calibration approach.

Li et al. (2018) performed a similar comparison with the MWA using the Fast Holographic Deconvolution (FHD) package (Sullivan et al. 2012) for sky-based calibration and the `omnical` package (Zheng et al. 2014) for redundant

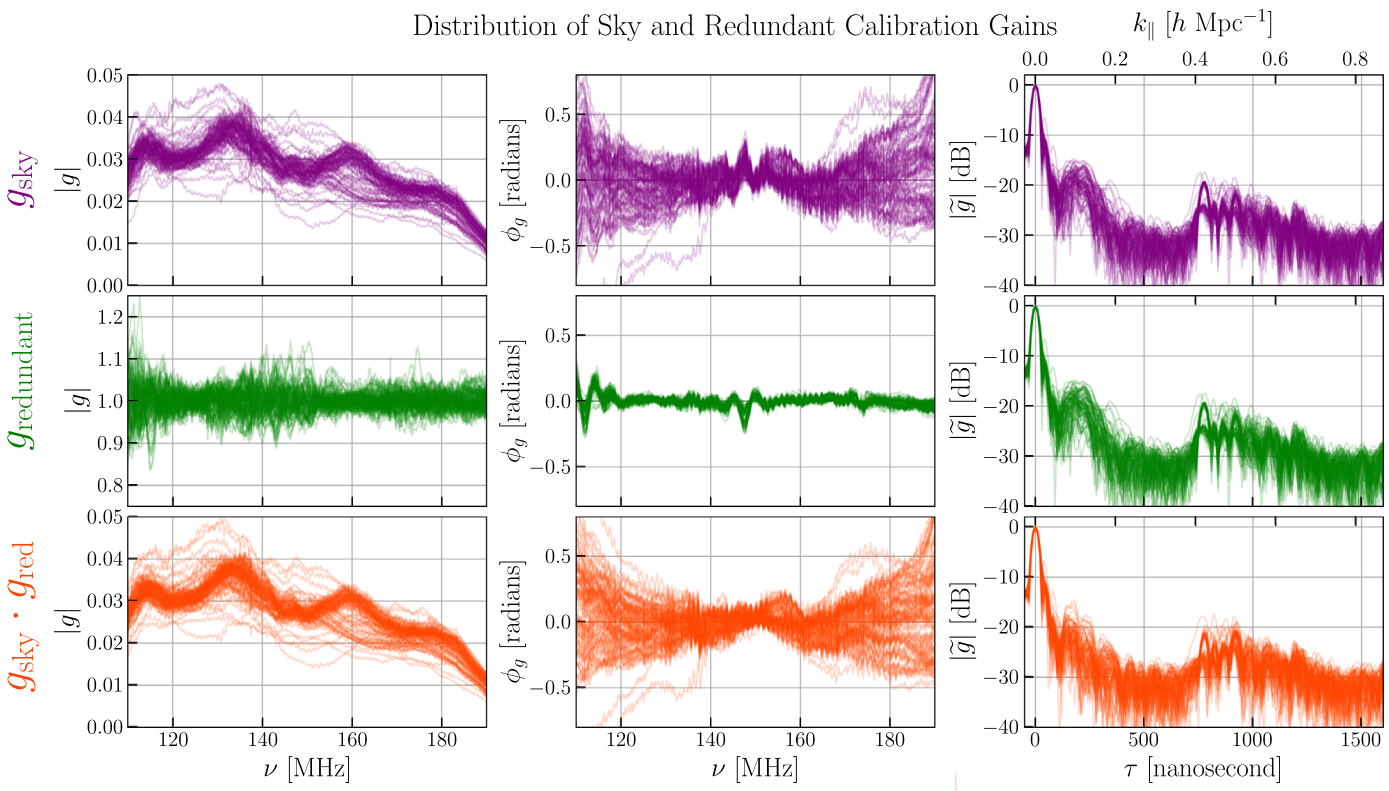

Figure 16. Distribution of gain solutions from AR calibration. The top panels show just the sky calibration (similar to Figure 10) in amplitude, phase (after removing their cable delay), and delay space. The middle panels show just the redundant calibration portion of the gains in AR calibration. The bottom panels show the product of the two steps, which forms the full AR calibration gain solution. Note the notch in the phase plots that is canceled out by redundant calibration, which leads to some suppression of the 200 ns feature.

calibration. Similar to this work, they found marginal improvements with a combined sky + redundant calibration approach.

6. Power Spectrum Performance

We use the visibility-based, delay spectrum estimator of the 21 cm power spectrum to further assess the quality of the calibration and the overall stability of the array. The delay transform is simply the Fourier transform of the visibilities across frequency into the delay domain

$$\widetilde{V}(\boldsymbol{u}, \tau) = \int d\nu \; e^{2\pi i \nu \tau} V(\boldsymbol{u}, \nu), \tag{8}$$

where $\boldsymbol{u} = \boldsymbol{b}/\lambda$ is the uv vector of the baseline and λ is the observing wavelength (Parsons et al. 2012a; Liu et al. 2014; Parsons et al. 2014). The Fourier dual of frequency, τ, is not a direct mapping of the line-of-sight spatial wavevector $k_\parallel$, but under certain assumptions, it is a fairly good approximation. This is known as the "delay approximation" and has been shown to be fairly accurate for short baselines (Parsons et al. 2012a). The delay spectrum estimate of the 21 cm power spectrum is the delay transformed visibilities squared, multiplied by the appropriate scaling factors,

$$\widehat{P}_{21}(k_\perp, k_\parallel) \approx |\widetilde{V}(\boldsymbol{u}, \tau)|^2 \frac{X^2 Y}{\Omega_{pp} B_p} \left(\frac{c^2}{2 k_B \bar{\nu}^2}\right)^2, \tag{9}$$

where X and Y convert angles on the sky and delay modes to cosmological length scales, Ω_{pp} is the sky integral of the squared primary beam, $\bar{\nu}$ is the average frequency in the delay transform window, and B_p is the delay transform bandwidth, as defined in Appendix B of Parsons et al. (2014). The relationships between the Fourier domains inherent to the telescope, $\boldsymbol{u}$ and τ, and the cosmological Fourier domains are

$$\begin{aligned} k_\parallel &= \frac{2\pi}{X}\tau \\ k_\perp &= \frac{2\pi}{Y}\frac{b}{\lambda}, \end{aligned} \tag{10}$$

where $X = c(1+z)^2 \nu_{21}^{-1} H(z)^{-1}$, $Y = D(z)$, $\nu_{21} = 1.420$ GHz, $H(z)$ is the Hubble parameter, $D(z)$ is the transverse comoving distance, b is the baseline length, and λ is the observing wavelength (Parsons et al. 2012b; Liu et al. 2014). For this analysis, we adopt a ΛCDM cosmology with parameters derived from the *Planck* 2015 analysis (Planck Collaboration et al. 2016), namely $\Omega_\Lambda = 0.6844$, $\Omega_b = 0.04911$, $\Omega_c = 0.26442$, and $H_0 = 67.27\,\mathrm{km\,s^{-1}}$ Mpc.

Due to the chromaticity of an interferometer, foreground emission that is inherently spectrally smooth (such as galactic synchrotron) will have increased spectral structure in the measured visibilities. The delay at which the instrument imparts this spectral structure is dependent on the geometric delay of the source signal between the two antennas that make up a baseline, given as

$$\tau = \frac{|\boldsymbol{b}|\sin(\theta)}{c}, \tag{11}$$

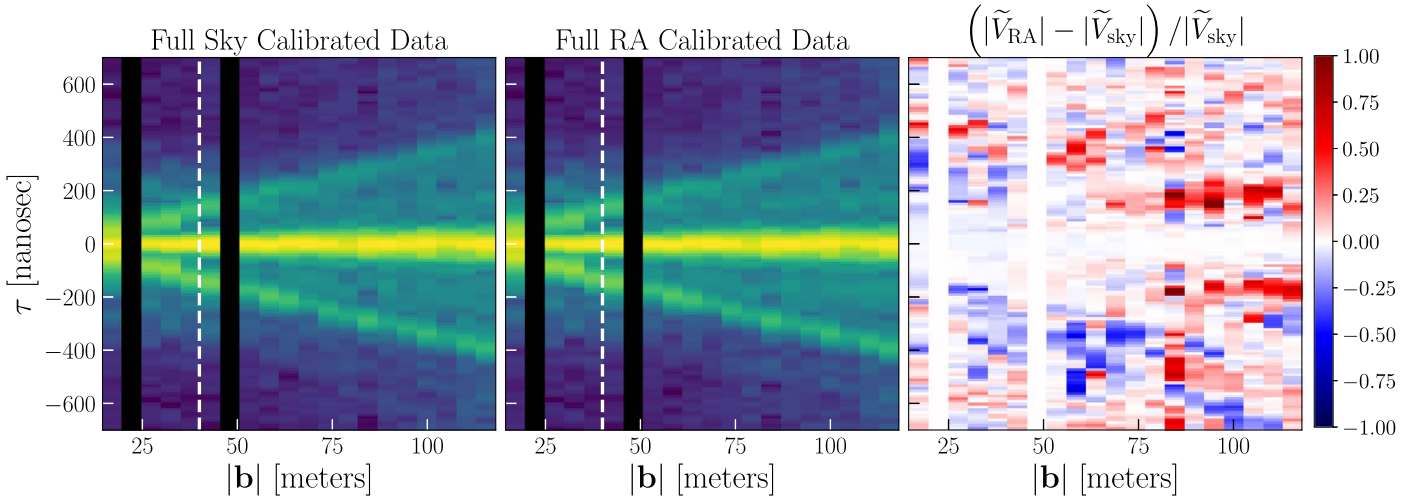

Figure 17. Redundantly averaged pseudo-Stokes *I* visibilities in delay space transformed across a 120–180 MHz spectral window and ordered according to baseline length, having applied the full sky (left) and RA (middle) calibration gains. These are plotted on the same color scale as Figure 9. Taking their fractional difference (right) shows that the RA calibration introduces less structure into the data at $|\tau| \sim 200$ ns for shorter baselines (blue regions), although it also seems to introduce new structure at slightly smaller delays for long baselines (red regions).

where θ is the zenith angle of the incident foreground emission and $\boldsymbol{b}$ is the baseline separation vector. We can see that spectrally smooth foregrounds incident from zenith will appear at lower delays and therefore have less induced chromaticity, while foregrounds incident from large zenith angles will have more induced chromaticity. The maximum delay at which a smooth-spectrum foreground can appear is called the horizon limit, in which case, $\tau_{\rm horizon} = \tau(\theta = 90°)$. If we could perfectly image the interferometric data, we could also reconstruct the smooth-spectrum foregrounds. However, this is never the case in practice, as effects like missing *uv* samples and imaging via gridded Fourier transforms create low-level chromatic side lobes that corrupt the images with spectrally dependent residual foregrounds. Visibility-based power spectrum estimators that do not even attempt to image the data are stuck with the most severe amounts of instrument-induced chromaticity, generally out to the baseline horizon delay. The horizon limit is a function of baseline length (Equation (11)), and as such, it forms a wedge-like shape in the data's Fourier domain and has come to be known as the foreground wedge (Datta et al. 2010; Morales et al. 2012; Parsons et al. 2012a; Thyagarajan et al. 2013; Liu et al. 2014; Morales et al. 2019). Because HERA has a fairly compact primary beam, we expect most foreground power to lie within $\tau \leqslant |\boldsymbol{b}| \sin(\theta = 5°)/c$; however, the vast amounts of diffuse emission near the horizon mean that we still expect to see some amount of foreground power out to the horizon limit, even though it is significantly attenuated by the primary beam (Thyagarajan et al. 2016).

The issue of whether foregrounds actually appear tightly confined within the foreground wedge is an open question: 21 cm foreground studies seem to indicate that suprahorizon foreground power tends to extend only slightly beyond the horizon (Bernardi et al. 2013; Pober et al. 2013; Gehlot et al. 2018; Lanman et al. 2019), but whether this is truly the case down to EoR sensitivities is not known. There are a number of effects that can contribute to measured suprahorizon foreground emission, including intrinsic foreground spectral structure, unflagged RFI, primary beam chromaticity, and gain calibration errors. As discussed in Section 4, the intrinsic gain kernel of the instrument may have a nonnegligible extent to large delay modes, which, if left uncalibrated, will push foreground power out to higher delays. Similarly, gain errors will introduce structure at these scales and have the same effect. Smoothing the gains eliminates the latter concern but still leaves the possibility of the former effect. To assess the degree of foreground containment, we can form wide-band, visibility-based power spectra as a diagnostic.

This is complemented by an understanding of how thermal noise appears in the power spectra. Given our knowledge of the noise properties of our antennas, we can compute a theoretical estimate of the noise power spectrum, $P_{\rm N}$, which is equivalent to the rms of the power spectrum if the only component in the data is noise. This is one measure of the uncertainty on the power spectra due to noise but also represents the theoretical amplitude of the power spectra in the limit that they are noise-dominated (as opposed to signal- or systematic-dominated). This is given in Cheng et al. (2018) as

$$P_{\rm N} = \frac{X^2 Y \Omega_{\rm eff} T_{\rm sys}^2}{t_{\rm int} N_{\rm coherent} \sqrt{2 N_{\rm incoherent}}}, \tag{12}$$

where the X and Y scalars are the same as before, $T_{\rm sys}$ is the system temperature in millikelvin, $t_{\rm int}$ is the correlator integration time in seconds, $N_{\rm coherent}$ is the number of sample averages done at the visibility level (i.e., before visibility squaring), and $N_{\rm incoherent}$ is the number of sample averages done at the power spectrum level (i.e., after visibility squaring). Here $\Omega_{\rm eff}$ is the effective beam area given by $\Omega_{\rm eff} = \Omega_p^2/\Omega_{pp}$, where Ω_p is the integral of the beam across the sky in steradians, and Ω_{pp} is the integral of the squared-beam across the sky in steradians (Pober et al. 2013; Parsons et al. 2014). Using similar data products, Kern et al. (2019a) showed that the HERA Phase I system achieves an antenna-averaged $T_{\rm sys} \sim 250$ K at 160 MHz, which we adopt in this work.

The raw data are flagged for RFI and thus nulled at the flagged channels. This leads to a highly discontinuous windowing function that, when taking a Fourier transform, will spread foreground power and contaminate the EoR

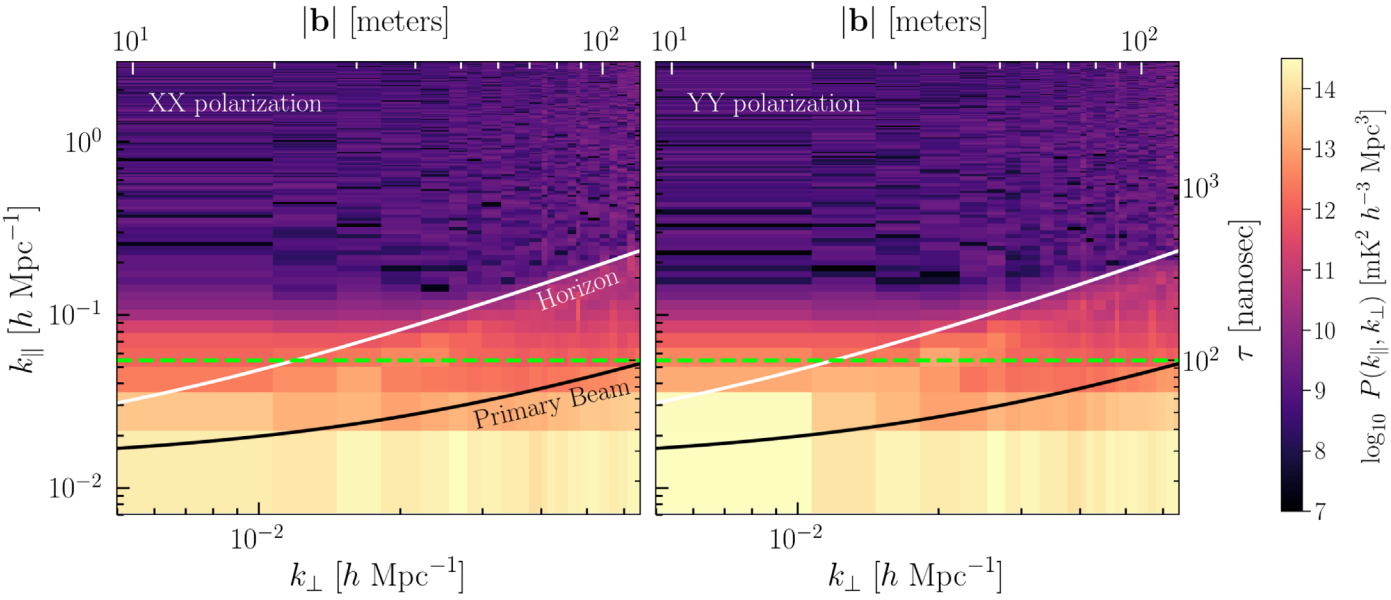

Figure 18. Wide-band, 2D power spectra of each linear dipole polarization *XX* (left) and *YY* (right), having applied the smooth sky-based calibration and after systematic removal and an incoherent average (i.e., after squaring the visibilities) from 0 to 2 hr LST. Power spectra are formed between 139 and 178 MHz, having applied a Blackman window to limit spectral leakage in the DFT. The black line marks the FWHM of the primary beam (±5° from zenith), and the white line marks the baseline horizon. Both lines have an additive buffer of $k_\parallel = 0.014\ h\ \mathrm{Mpc}^{-1}$ to account for the width of the Blackman kernel in Fourier space. The dashed green line marks the maximum delay scale of the smoothed gain solutions. Most of the foreground power is confined within the horizon limit of the array; however, there is evidence for some suprahorizon leakage at short baselines.

window. To prevent this, we employ the same 1D delay domain deconvolution as the gain-smoothing filter (Section 4) on each visibility, filling in model CLEAN components out to 2000 ns. HERA Phase I data are contaminated by cable reflection and cross-coupling instrumental systematics (Kern et al. 2019a). Because we are concerned with foreground leakage due to calibration in this work, we remove these systematics before forming power spectra for visual clarity. Specifically, we apply a time-domain filter to suppress cross-coupling in all visibilities with a projected east–west length greater than 14 m (throwing out all other visibilities). This time-domain filter is performed in the delay domain and isolates a rectangle spanning $|\tau| > 0.8\tau_{\mathrm{horizon}}$ and fringe rates given by the 99% EoR power bounds in Kern et al. (2019b) for each baseline independently. We also calibrate out a single cable reflection term for each of the 20 and 150 m cables in the analog system per dipole polarization, also using the methods in Kern et al. (2019b).

We form power spectra across a spectral window from 139 to 178 MHz and apply a Blackman window prior to taking the Fourier transform to limit spectral leakage in the DFT. Baselines are only cross-multiplied with themselves and not with other baselines in a redundant group. Normally, this would produce a noise bias in the power spectra, so instead, we cross-multiply baselines with themselves at adjacent time integrations, having first rephased them to the same pointing center (Pober et al. 2013). We do this for all baselines for each time integration pair in the range of 0–2 hr LST. After squaring the visibilities, we incoherently average the power spectra across LST and then average all power spectra of the same baseline length (regardless of orientation), which is equivalent to cylindrically gridding $\boldsymbol{k}$ space into $k_\perp$and $k_\parallel$ annuli.

Figure 18 shows the 2D power spectra in instrumental *XX* and *YY* visibility polarizations with the smooth sky calibration gains applied to the data (smoothed out to $\tau = 100$ ns). We also show the primary beam FWHM limit (black) and full horizon limit (white) in both instrumental *XX* and *YY* visibility polarization. Both lines have an additive buffer of $k_\parallel = 0.014\ h\ \mathrm{Mpc}^{-1}$ to account for the width of the Blackman kernel in Fourier space. The dashed green line shows the maximum delay scale of the applied gains after smoothing. We find that most of the foreground power is contained within the horizon limit, with some amounts of suprahorizon leakage for short baselines. The strong pitchfork feature of the foreground emission tracing the horizon line is not as prominent in this plot as it was in Figure 9, which is due to the fact that it was partially removed in the cross-coupling filter applied to the data. Kern et al. (2019a) showed that the edges of the pitchfork are slowly time variable and thus can be separated from the cosmological 21 cm signal and filtered out with a high-pass time filter. Figure 11 also demonstrates this, showing that the two lobes at $\pm\tau = 100$ ns are also centered at $f = 0$ mHz, meaning they primarily contain slowly time-variable terms. This means that the time filter designed to eliminate cross-coupling also helps to reduce some of the strongest foreground emission straddling the boundary of the foreground wedge and EoR window.

To first order, Figure 18 tells us that our single-field, smoothed sky-based calibration with restricted degrees of freedom has done a fairly good job calibrating the data and largely kept foreground power contained within the foreground wedge. For short baselines, however, we can begin to see evidence for some amount of suprahorizon emission that could be due to uncalibrated gain terms or imperfectly removed cross-coupling, the latter of which is harder to remove for shorter baselines. This suprahorizon emission is located beyond the smoothing scale of the gains and appears in amplitude slightly larger than predictions of the high-order dish reflections (Patra et al. 2018) but is contained within $k_\parallel \lesssim 0.2\ h\ \mathrm{Mpc}^{-1}$ down to nearly $\sim 10^6$ in dynamic range against the foreground peak. Note that, possibly coincidentally, this suprahorizon excess seems more prevalent for baselines shorter

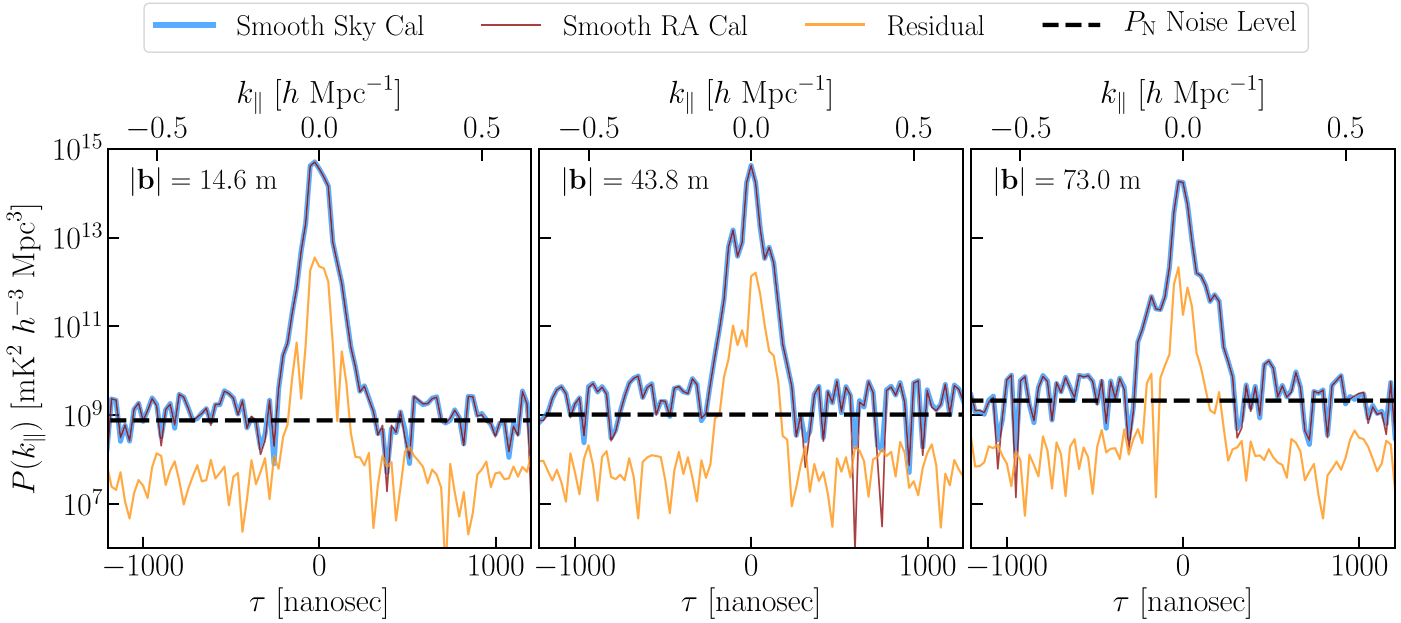

Figure 19. Delay spectra of three redundantly averaged east–west baseline types for the instrumental *YY* polarization, showing the data calibrated with the smooth sky (blue) and RA (red) calibrations and their residual (orange), along with the thermal noise floor (dashed black line) assuming a $T_{\rm sys} = 250$ K. The two calibrations yield nearly the same averaged power spectra across all delays, which shows consistency with the theoretical noise floor outside $k_\parallel \gtrsim 0.2\ h$ Mpc^{-1}.

than our initial baseline cut of 40 m. Deeper integrations will help to discriminate whether the observed suprahorizon emission extends further out in $k_\parallel$ space at lower noise levels.

Figure 19 shows the same power spectra but focuses on three unique baseline types: purely east–west baselines of 14.6, 43.8, and 73 m in length. In addition to showing the power spectra of the data with the smooth sky calibration (blue), we also show the smooth RA calibration (also smoothed out to $\tau = 100$ ns) and the residual between the two. This demonstrates that the calibration strategies, postsmoothing, have nearly the same impact on the averaged power spectrum. We also show the theoretical noise floor of the data (dashed black line), which more clearly demonstrates the agreement of the data with the noise floor outside $k_\parallel \gtrsim 0.2\,h\,{\rm Mpc}^{-1}$. Note that the noise floor for longer baseline types is higher because there are fewer physical baselines, meaning less averaging is done in the (coherent) redundant average.

7. Summary

In this work, we discuss sky-based and hybrid-redundant calibration strategies for Phase I of HERA. We present a CASA-based calibration pipeline for constructing sky models, applying primary beam corrections, and performing direction-independent, antenna-based complex gain calibration. We use this to characterize the time and frequency stability of the gain solutions, finding that slow and fast nightly time drifts in the gain's overall flux scale are of order 8% and 0.1% effects, respectively. We also show that unmodeled diffuse foregrounds, mutual coupling, and other cross-coupling systematics in the HERA Phase I system are picked up in the process of both sky and redundant calibration, which introduces errors in gain calibration at delay scales $\tau > 100$ ns. By low-pass filtering the gains with a Fourier filter, we can restrict the degrees of freedom of the gains, and we show that this mitigates the effect of these gain errors. Additionally, while we do not perform any kind of polarization calibration, we find that polarization leakage from Stokes I to Q, U, and V is on the order of a few percent.

We also present a hybrid approach for combining redundant and absolute calibration techniques and, similar to Li et al. (2018), find that the hybrid technique marginally improves the gain solutions over just sky-based calibration, although we did not perform any self-calibration iterations, which would likely have improved the fidelity of the sky-based gains. This was omitted because of the difficulty of implementing this to a high dynamic range given the mediocre Phase I angular resolution but will be explored in future work. Additionally, we show that the hybrid calibration scheme is also limited by gain errors at similar $\tau \gtrsim 100$ ns scales as the sky calibration, which we suggest can be further mitigated by enacting a larger minimum baseline cut in the calibration, as well as attempts to include the diffuse emission component of the sky into the calibration model.

Finally, we form 2D power spectra across 139–178 MHz from 0 to 2 hr LST and show that most of the foreground power measured by HERA is contained within the horizon limit of the array, but we do observe nonnegligible suprahorizon power for short baselines that were not included in the calibration. This emission is confined within $k_\parallel \lesssim 0.2\,h\,{\rm Mpc}^{-1}$ down to the noise floor of the data, which achieves a dynamic range of nearly 10^6 against the peak foreground power. This could be due to uncalibrated gain terms at these scales or residual instrumental systematics. Deeper integrations will help make this clearer and help us understand how far in $k_\parallel$ the suprahorizon emission extends. In repeating the analysis for both the smooth sky calibration and smooth hybrid calibration, we find that they have nearly the same impact on the power spectra. Future observing seasons with the full HERA array will make high dynamic range imaging and direction-dependent calibration easier to implement and may be a way to mitigate some of the errors observed in the gain solutions. Overall, our work shows that HERA Phase I can be relatively well calibrated for a foreground avoidance power spectrum

estimator with only a few degrees of freedom across the time and frequency axes.

This material is based upon work supported by the National Science Foundation under grant Nos. 1636646 and 1836019 and institutional support from the HERA collaboration partners. This research is funded in part by the Gordon and Betty Moore Foundation. HERA is hosted by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. J.S.D. gratefully acknowledges the support of the NSF AAPF award No. 1701536. A.L. acknowledges support from a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and a Discovery Launch Supplement, as well as the Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholars program. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in three-Dimensions (ASTRO 3D) through project No. CE170100013. G.B. acknowledges funding from the INAF PRIN-SKA 2017 project 1.05.01.88.04 (FORE-CaST), support from the Ministero degli Affari Esteri della Cooperazione Internazionale—Direzione Generale per la Promozione del Sistema Paese Progetto di Grande Rilevanza ZA18GR02 and the National Research Foundation of South Africa (grant No. 113121) as part of the ISARP RADIO-SKY2020 Joint Research Scheme, from the Royal Society and the Newton Fund under grant NA150184, and from the National Research Foundation of South Africa (grant No. 103424).

Appendix A
Partial Absolute Calibration

Partial absolute calibration is the process of taking a set of sky model visibilities and setting up a system of equations that solves for just the degenerate components of redundant calibration. The number of degenerate modes in redundant calibration depends on the kind of redundant calibration being employed (Dillon et al. 2018). Here we discuss the degeneracies associated with the "2-pol" scheme, which calibrates the X and Y dipoles separately and ignores cross-feed polarization terms. As shown in Section 5, there are three main degeneracies in redundant calibration for each dipole polarization: the average gain amplitude (or the absolute flux scale of the instrument) and a "tip-tilt" phase gradient as a function of distance from the center of the array for both the east and north spatial axes (or the overall pointing center of the instrument). Each of these parameters has an arbitrary frequency dependence, meaning that various kinds of spectral structure can occupy these degenerate modes. We can express these parameters in the ith antenna gain of the X dipole as

$$g_{i,X}(\nu) = \exp(\eta_{\mathrm{abs},X}(\nu) + i2\pi\nu(T_{E,X}r_{i,E} + T_{N,X}r_{i,N}) + i(\Phi_{E,X}(\nu)r_{i,E} + \Phi_{N,X}(\nu)r_{i,N})), \quad (13)$$

where $r_{i,E}$ is the east distance of antenna i from the center of the array in meters, and we have explicitly included the frequency dependence of the gain and its parameters. Note that we have redefined the phase component into the sum of two terms, a spatial delay gradient $\boldsymbol{T}_X = (T_{E,X}, T_{N,X})$ and a spatial phase gradient $\boldsymbol{\Phi}_X = (\Phi_{E,X}, \Phi_{N,X})$. Note that the delay gradient parameter has units of nanoseconds per meter and is itself frequency-independent, but it has the effect of creating a phase slope in the gain across frequency. The delay gradient manifests as a delay plane across the array that sets the phase center. It forms a subspace of the original phase gradient space, so we simply pull it out and redefine the phase gradient term Φ to be a deviation about the delay plane. This is important because when we go to solve the calibration equation, we want the phase measurements to be near zero, or at least considerably less than 2π, to mitigate phase wrapping (Liu et al. 2010). Phase wrapping creates local minima that confuse the calibration phase solver, which can be alleviated through preconditioning of the system of equations by first solving for and eliminating the delay gradient term.

Using a logarithm to linearize the calibration equation, the average antenna amplitude for the X dipole is found by solving the following system of equations:

$$\boldsymbol{y} = \begin{pmatrix} \ln\left|\frac{V^{\mathrm{data}}_{ij,XX}}{V^{\mathrm{model}}_{ij,XX}}\right| \\ \ln\left|\frac{V^{\mathrm{data}}_{jk,XX}}{V^{\mathrm{model}}_{jk,XX}}\right| \\ \vdots \end{pmatrix} = \boldsymbol{A}\hat{\boldsymbol{x}} = \begin{pmatrix} 2 \\ 2 \\ \vdots \end{pmatrix}\left(\hat{\eta}_{\mathrm{abs},X}\right), \quad (14)$$

where we have specified now that the visibilities are from the XX instrumental polarization. We use the linear and nonlinear equation-solving package `linsolve` to solve these systems of equations.

The delay gradient parameter can be isolated by taking the phase of the data–model ratio,

$$\mathrm{angle}\left(\frac{V^{\mathrm{data}}_{ij,XX}}{V^{\mathrm{model}}_{ij,XX}}\right)(\nu) = 2\pi\nu\boldsymbol{T}_X\boldsymbol{r}_{ij}, \quad (15)$$

where the angle $(\times)$ operator is $\tan^{-1}(\mathrm{Im}(\times)/\mathrm{Re}(\times))$. However, we can see that the delay gradient parameter is not inherently a function of frequency, so instead of solving this equation at each frequency, we should recast it in a form that is frequency-independent and then solve that equation. This can be expressed as

$$\mathrm{delay}\left(\frac{V^{\mathrm{data}}_{ij,XX}}{V^{\mathrm{model}}_{ij,XX}}\right) = \boldsymbol{T}_X\boldsymbol{r}_{ij}, \quad (16)$$

where the delay $(\times)$ operator takes the Fourier transform of its argument and identifies the delay of its peak in amplitude via quadratic interpolation of the three strongest Fourier modes. The system of equations for the delay gradient of the X dipole is then

$$\boldsymbol{y} = \begin{pmatrix} \mathrm{delay}\left(\frac{V^{\mathrm{data}}_{ij,XX}}{V^{\mathrm{model}}_{ij,XX}}\right) \\ \mathrm{delay}\left(\frac{V^{\mathrm{data}}_{jk,XX}}{V^{\mathrm{model}}_{jk,XX}}\right) \\ \vdots \end{pmatrix} = \boldsymbol{A}\hat{\boldsymbol{x}} = \begin{pmatrix} r_{ij,E} & r_{ij,N} \\ r_{jk,E} & r_{jk,N} \\ \vdots & \end{pmatrix}\begin{pmatrix} \hat{T}_{E,X} \\ \hat{T}_{N,X} \end{pmatrix}. \quad (17)$$

The estimated delay gradient gain is then expressed as $\hat{g}_{i,X}(\nu) = \exp(i2\pi\nu\hat{\boldsymbol{T}}_X\boldsymbol{r}_i)$.

After dividing the data visibilities by the estimated delay gradient gains, we can solve for the leftover phase gradient

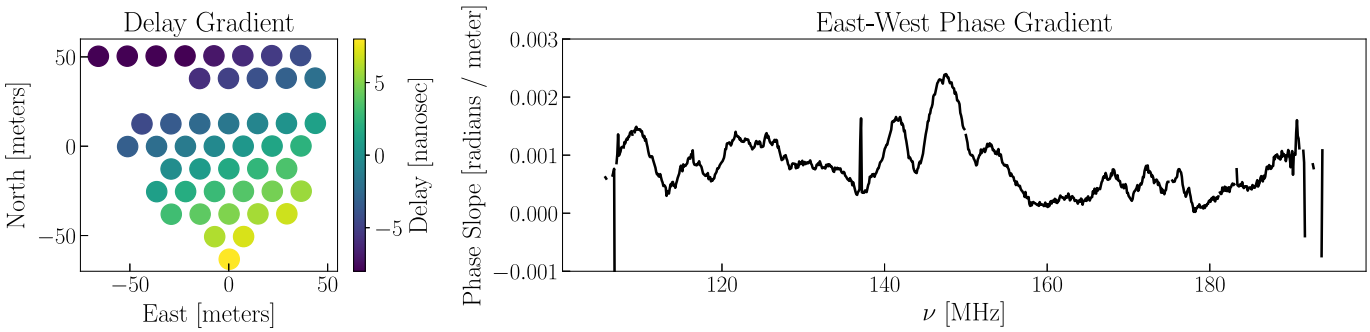

Figure 20. Delay (left) and east–west phase (right) gradients derived by partial absolute calibration for the *X*-dipole polarization using the GLEAM-02H field flux density model. We observe a significant amount of spectral structure in the phase gradient parameter, meaning it cannot be overlooked in partial absolute calibration.

parameter for the *X* dipole with the system of equations

$$\boldsymbol{y} = \begin{pmatrix} \text{angle}\left(\frac{\widetilde{V}^{\text{data}}_{ij,XX}}{V^{\text{model}}_{ij,XX}}\right) \\ \text{angle}\left(\frac{\widetilde{V}^{\text{data}}_{jk,XX}}{V^{\text{model}}_{jk,XX}}\right) \\ \vdots \end{pmatrix} = \boldsymbol{A}\hat{\boldsymbol{x}} = \begin{pmatrix} r_{ij,E} & r_{ij,N} \\ r_{jk,E} & r_{jk,N} \\ \vdots & \end{pmatrix} \begin{pmatrix} \hat{\Phi}_{E,X} \\ \hat{\Phi}_{N,X} \end{pmatrix}, \quad (18)$$

where $\widetilde{V}^{\text{data}}$ denotes the fact that we have first divided the data visibilities by the delay gradient gain (or, equivalently, multiplied the model visibilities by the delay gradient gain). The average amplitude parameter, being orthogonal to the phase parameters, does not necessarily need to precede these steps. The full partial absolute calibration gain is then simply Equation (13) filled with our estimates of the degenerate parameters.

One interesting feature about delay and phase gradient calibration is that it does not require a reference antenna. Because phase is a periodic coordinate system, sky-based phase calibration requires that we select a reference antenna whose phase is identically zero, which is a way to constrain the overall phase parameter that does not have a physical meaning. If the array coordinates are defined in east–north–up coordinates, then for delay and phase gradient calibration, we can change the overall phase of the gain solutions by moving the estimated delay or phase plane up or down along the up-axis (i.e., the *z*-axis if the array is defined in *X*–*Y*–*Z* coordinates). Moving the delay and phase planes up or down does not change the relative delay and phase between antennas, which is really what we care about. We can pin this free parameter by selecting the up-axis intercept of the plane, which is equivalent to setting the origin of the $\boldsymbol{r}_i$ vector coordinates that, up to now, we have defined as the center of the array. We can see now that setting it at the center of the array is not strictly required, but it does make computations easier. Therefore, the act of setting the origin of the antenna position coordinate system plays the same role in delay and phase gradient calibration as the reference antenna does in standard sky-based phase calibration.

In Figure 20, we show the derived delay gradient (left) and phase gradient (right) parameters across the array for the *X*-dipole polarization. The phase gradient term shows significant amounts of spectral structure, highlighting its ability to pick up on nontrivial spectral terms in the data. For a large, 350+ element array, these steps may take too long to calibrate the data in real time using all $\sim N^2$ baselines. However, for partial absolute calibration, we may get away with only using some of the baselines instead of all of them. The degenerate parameters of redundant calibration outlined above are not actually specific to any individual antenna in the array; they are only properties of the array itself. We could, for example, use only a fraction of the longest baselines in the array for partial absolute calibration, which gives us a lever-arm advantage for estimating the delay and phase gradient terms. Concerns about this approach are (1) increased noise in the gains due to less data points in our $\boldsymbol{y}$ vector and (2) whether the baselines selected are drawn from a unique population of antennas relative to all other antennas in the array, in which case, the average amplitude and phase gradients estimated with the longest baselines (which will preferably come from antennas near the edge of the array) will be misestimates for the other antennas not represented in the system of equations. One could devise strategies for mitigating these kinds of concerns by, say, ensuring that while only a fraction of the baselines are used in calibration, every antenna is at least somewhat represented in the system of equations.

Appendix B
Software

The analysis presented in this work relies heavily on the Python programming language (https://www.python.org) and Python software developed by HERA collaboration members. Here we provide a list of these packages and their versions: `aipy [v2.1.12]` (https://github.com/HERA-Team/aipy), `healvis [v1.0.0]` (https://github.com/RadioAstronomySoftwareGroup/healvis; Lanman & Kern 2019), `hera_cal [v2.0]` (https://github.com/HERA-Team/hera_cal), `hera_sim [v0.0.1]` (https://github.com/HERA-Team/hera_sim), `pyuvdata [v1.3.6]` (https://github.com/RadioAstronomySoftwareGroup/pyuvdata; Hazelton et al. 2017), and `uvtools [v0.1.0]` (https://github.com/HERA-Team/uvtools). These packages, in turn, rely heavily on other publicly available software packages, including `astropy [v2.0.14]` (https://astropy.org; The Astropy Collaboration et al. 2013), `healpy [v1.12.9]` (https://github.com/healpy/healpy), `h5py [v2.8.0]` (https://www.h5py.org/), `matplotlib [v2.2.4]` (https://matplotlib.org), `numpy [v1.16.2]` (https://www.numpy.org), `scipy [v1.2.1]` (https://scipy.org), and `scikit-learn [v0.19.2]` (https://scikit-learn.org).

ORCID iDs

Nicholas S. Kern https://orcid.org/0000-0002-8211-1892
Joshua S. Dillon https://orcid.org/0000-0003-3336-9958

Christopher L. Carilli https://orcid.org/0000-0001-6647-3861
Gianni Bernardi https://orcid.org/0000-0002-0916-7443
James E. Aguirre https://orcid.org/0000-0002-4810-666X
Adam P. Beardsley https://orcid.org/0000-0001-9428-8233
Judd D. Bowman https://orcid.org/0000-0002-8475-2036
Philip Bull https://orcid.org/0000-0001-5668-3101
John Ely https://orcid.org/0000-0002-4814-5511
Aaron Ewall-Wice https://orcid.org/0000-0002-0086-7363
Steve R. Furlanetto https://orcid.org/0000-0002-0658-1243
Bryna J. Hazelton https://orcid.org/0000-0001-7532-645X
Daniel C. Jacobs https://orcid.org/0000-0002-0917-2269
Joshua Kerrigan https://orcid.org/0000-0002-1876-272X
Piyanat Kittiwisit https://orcid.org/0000-0003-0953-313X
Saul A. Kohn https://orcid.org/0000-0001-6744-5328
Matthew Kolopanis https://orcid.org/0000-0002-2950-2974
Paul La Plante https://orcid.org/0000-0002-4693-0102
Adrian Liu https://orcid.org/0000-0001-6876-0928
Miguel F. Morales https://orcid.org/0000-0001-7694-4030
Steven G. Murray https://orcid.org/0000-0003-3059-3823
Abraham R. Neben https://orcid.org/0000-0001-7776-7240
Chuneeta D. Nunhokee https://orcid.org/0000-0002-5445-6586
Nipanjana Patra https://orcid.org/0000-0002-9457-1941
Jonathan C. Pober https://orcid.org/0000-0002-3492-0433
Nithyanandan Thyagarajan https://orcid.org/0000-0003-1602-7868
Peter K. G. Williams https://orcid.org/0000-0003-3734-3587

References

Barry, N., Beardsley, A. P., Byrne, R., et al. 2019a, PASA, 36, e026
Barry, N., Hazelton, B., Sullivan, I., Morales, M., & Pober, J. 2016, MNRAS, 461, 3135
Barry, N., Wilensky, M., Trott, C.M., et al. 2019b, ApJ, 884, 1
Beardsley, A., Hazelton, B., Sullivan, I., et al. 2016, ApJ, 833, 102
Bernardi, G., Greenhill, L.J., Mitchell, D.A., et al. 2013, ApJ, 771, 105
Bernardi, G., Zwart, J.T.L., Price, D., et al. 2016, MNRAS, 461, 2847
Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Blackman, R. B., & Tukey, J. W. 1958, BSTJ, 37, 185
Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, Natur, 555, 67
Byrne, R., Morales, M.F., Hazelton, B., et al. 2019, ApJ, 875, 70
Carilli, C. L., Nikolic, B., Thyagarayan, N., & Gale-Sides, K. 2018, RaSc, 53, 845
Cheng, C., Parsons, A.R., Kolopanis, M., et al. 2018, ApJ, 868, 26
Condon, J. J. 1997, PASP, 109, 166
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, ISTSP, 2, 647
Datta, A., Bowman, J., & Carilli, C. 2010, ApJ, 724, 526
de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, MNRAS, 388, 247
DeBoer, D., Siemion, A.P.V., Marcy, G.W., et al. 2017, PASP, 129, 45001
Dillon, J., & Parsons, A. 2016, ApJ, 826, 181
Dillon, J. S., Kohn, S.A., Parsons, A.R., et al. 2018, MNRAS, 477, 5670
Dillon, J., Liu, A., Williams, C., et al. 2014, PhRvD, 89, 23002
Eastwood, M. W., Anderson, M.M., Monroe, R.M., et al. 2018, AJ, 156, 32
Eastwood, M. W., Anderson, M.M., Monroe, R.M., et al. 2019, AJ, 158, 84
Ewall-Wice, A., Dillon, J. S., Liu, A., & Hewitt, J. 2017, MNRAS, 470, 1849
Ewall-Wice, A., Dillon, J., Hewitt, J., et al. 2016a, MNRAS, 460, 4320
Ewall-Wice, A., Bradley, R., DeBoer, D., et al. 2016b, ApJ, 831, 196
Fagnoni, N., de Lera Acedo, E., DeBoer, D.R., et al. 2019, arXiv:1908.02383
Gehlot, B. K., Koopmans, L.V.E., de Bruyn, A.G., et al. 2018, MNRAS, 478, 1484
Gehlot, B. K., Mertens, F.G., Koopmans, L.V.E., et al. 2019, MNRAS, 488, 4271
Hamaker, J. P., & Bregman, J. D. 1996, A&As, 117, 161
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&As, 117, 137
Hazelton, B. J., Jacobs, D. C., Pober, J. C., & Beardsley, A. P. 2017, JOSS, 2, 140
Högbom, J. A. 1974, A&As, 15, 417
Hurley-Walker, N., Callingham, J.R., Hancock, P.J., et al. 2017, MNRAS, 464, 1146
Intema, H. T., van der Tol, S., & Cotton, W.D. 2014, A&A, 501, 1185
Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78
Jacobs, D. 2016, HERA Calibrator Sources, Tech. Rep., School of Earth and Space Exploration, Arizona State University, Tempe, AZ, HERA Memo #17
Jacobs, D. C., Parsons, A.R., Aguirre, J.E., et al. 2013, ApJ, 776, 108
Jacobs, D., Pober, J., Parsons, A., et al. 2015, ApJ, 801, 51
Joseph, R. C., Trott, C. M., & Wayth, R. B. 2018, AJ, 156, 285
Kern, N., Parsons, A. R., Dillon, J. S., et al. 2019, ApJ, 884, 105
Kern, N., Parsons, A. R., Dillon, J. S., & Lanman, A. 2020, ApJ, 888, 70
Kerrigan, J. R., Pober, J.C., Ali, Z.S., et al. 2018, ApJ, 864, 131
Kerrigan, J., La Plante, P., Kohn, S., et al. 2019, arXiv:1902.08244
Kohn, S. A., Aguirre, J.E., La Plante, P., et al. 2019a, ApJ, 882, 58
Kohn, S., Aguirre, J., Nunhokee, C., et al. 2016, ApJ, 823, 88
Kolopanis, M., Jacobs, D.C., Cheng, C,, et al. 2019, ApJ, 883, 133
Lane, W. M., Cotton, W. D., van Velzen, S., et al. 2014, MNRAS, 440, 327
Lanman, A. E., & Kern, N. 2019, Healvis: Radio Interferometric Visibility Simulator based on HEALpix maps, Astrophysics Source Code Library
Lanman, A. E., Pober, J. C., Kern, N. S., et al. 2019, MNRAS, 487, 5840
Large, M. I., Mills, B. Y., Little, A. G., Crawford, D. F., & Sutton, J. M. 1981, MNRAS, 194, 693
Lenc, E., Anderson, C.S., Barry, N., et al. 2017, PASA, 34, e040
Li, W., Pober, J.C., Hazelton, B.J., et al. 2018, ApJ, 863, 170
Li, W., Pober, J.C., Barry, N., et al. 2019, arXiv:1911.10216
Liu, A., Parsons, A., & Trott, C. 2014, PhRvD, 90, 23018
Liu, A., & Shaw, J. R. 2019, arXiv:1907.08211
Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., & Zaldarriaga, M. 2010, MNRAS, 408, 1029
Martinot, Z. E., Aguirre, J. E., Kohn, S. A., & Washington, I. Q. 2018, ApJ, 869, 79
McKinley, B., Yang, R., Lopez-Caniego, M., et al. 2015, MNRAS, 446, 3478
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in ASP Conf. Ser. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 127
Mesinger, A. 2016, Understanding the Epoch of Cosmic Reionization (Cham: Springer), 423
Moore, D. F., Aguirre, J.E., Kohn, S.A., et al. 2017, ApJ, 836, 154
Morales, M., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
Morales, M., & Wyithe, J. 2010, ARA&A, 48, 127
Morales, M. F., Beardsley, A., Pober, J., et al. 2019, MNRAS, 483, 2207
Neben, A. R., Bradley, R.F., Hewitt, J.N., et al. 2016, ApJ, 826, 199
Nunhokee, C. D., Bernardi, G., Kohn, S.A., et al. 2017, ApJ, 848, 47
Orosz, N., Dillon, J. S., Ewall-Wice, A., Parsons, A. R., & Thyagarajan, N. 2019, MNRAS, 487, 537
Paciga, G., Albert, J., Bandura, K., et al. 2013, MNRAS, 433, 639
Parsons, A., Pober, J., Aguirre, J., et al. 2012a, ApJ, 756, 165
Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012b, ApJ, 753, 81
Parsons, A. R., Liu, A., Ali, Z. S., & Cheng, C. 2016, ApJ, 820, 51
Parsons, A. R., Backer, D.C., Foster, G.S., et al. 2010, AJ, 139, 1468
Parsons, A., Liu, A., Aguirre, J., et al. 2014, ApJ, 788, 106
Patil, A., Yatawatta, S., Koopmans, L., et al. 2017, ApJ, 838, 65
Patra, N., Parsons, A.R., DeBoer, D.R., et al. 2018, ExA, 45, 177
Planck Collaboration et al. 2016, A&A, 594, A13
Pober, J. C., Parsons, A.R., Jacobs, D.C., et al. 2012, AJ, 143, 53
Pober, J., Parsons, A., Aguirre, J., et al. 2013, ApJL, 768, L36
Sault, R. J., Hamaker, J. P., & Bregman, J. D. 1996, A&As, 117, 149
Singh, S., Subrahmanyan, R., Udaya Shankar, N., et al. 2017, arXiv:1703.06647
Sullivan, I. S., Morales, M.F., Hazelton, B.J., et al. 2012, ApJ, 759, 17
The Astropy Collaboration et al. 2013, A&A, 558, 558
Thyagarajan, N., Parsons, A. R., DeBoer, D. R., et al. 2016, ApJ, 825, 9
Thyagarajan, N., Udaya Shankar, N., Subrahmanyan, R., et al. 2013, ApJ, 776, 6
Thyagarajan, N., Jacobs, D., Bowman, J., et al. 2015, ApJ, 804, 14
Tingay, S. J., Goeke, R., Bowman, J.D., et al. 2013, PASA, 30, e007
van Haarlem, M. P., Wise, M.W., Gunst, A.W., et al. 2013, A&A, 556, A2
Wieringa, M. H. 1992, ExA, 2, 203
Zheng, H., Tegmark, M., Buza, V., et al. 2014, MNRAS, 445, 1084
Zheng, H., Tegmark, M., Dillon, J.S., et al. 2017, MNRAS, 464, 3486